Fall 12-20-2020

Optical Signatures of Plankton in the Open Ocean: From Individual Cells to Global Patterns

Nils Haentjens
University of Maine, nils.haentjens@maine.edu

OPTICAL SIGNATURES OF PLANKTON IN THE OPEN OCEAN: FROM INDIVIDUAL CELLS TO GLOBAL PATTERNS

By

Nils Haëntjens

M.E., Institut Supérieur d'Électronique de Paris, 2015

M.S., University of Maine, 2017

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Oceanography)

The Graduate School

The University of Maine

December 2020

Advisory Committee:

Emmanuel Boss, Ph.D., Professor of Oceanography, Advisor

Lee Karp-Boss, Ph.D., Associate Professor of Oceanography

Andrew Thomas, Ph.D., Professor of Oceanography

Jason Graff, Ph.D., Assistant Professor of Oceanography

Fei Chai, Ph.D., Professor of Oceanography

OPTICAL SIGNATURES OF PLANKTON IN THE OPEN OCEAN: FROM INDIVIDUAL CELLS TO GLOBAL PATTERNS

By Nils Haëntjens

Dissertation Advisor: Dr. Emmanuel Boss

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Oceanography)
December 2020

Marine plankton ecosystems play a major role on Earth, having implications for the global carbon cycle and the food-web structures. Ocean color satellites and networks of autonomous platforms equipped with optical sensors are the primary tools used to study phytoplankton dynamics. They provide long term records while offering a synoptic view of our oceans, enabling to study impact of climate variability on planktonic ecosystems. Interpretation of these observations rely heavily on optical theory and how light propagating through the water is affected by particles who absorb and scatter light (e.g. phytoplankton, sediments). However, the complexity of the optical properties of natural seawater often obscures their interpretation.

I address some of the current challenges in optical theory by analyzing measurements of inherent optical properties and phytoplankton size distribution (PSD). The PSD built spans four seasons across regions of the western North Atlantic, including large variability which highlight the dynamic annual cycle of phytoplankton of this area. Previously established algorithms used to estimate phytoplankton size algorithms based of optical properties are assessed as to date they have not been validated with actual size measurements. Additionally, the contribution of phytoplankton to particulate attenuation and backscattering and its efficiency to absorb light are computed for the upper ocean.

The PSDs revealed that phytoplankton dominate attenuation and backscattering signals ($\sim$ 75 %) reinforcing the idea that these properties are good predictors of phytoplankton biomass. Additionally, spectral slopes of attenuation and backscattering also correlate well with the PSD. This suggests that ocean color algorithms should focus on improved retrieval of backscattering spectra.

A data logger was developed to improve current recording of optical data during long term deployment on research vessels. It was successfully deployed >650 days at sea. Finally, I proposed a novel method to detect a subset of diel migrating organisms (SDMO) responsible for anomalies in particulate backscattering and ultra-violet fluorescent profiles from autonomous platforms. This method demonstrates the seasonality of SDMO in the world's ocean in regions where such data has been lacking and provides the distribution of SDMO which play an important role in the biological pump, extending observations beyond "classical" methodology.

ACKNOWLEDGEMENTS

This thesis, as with any project that spans more than two oceans and five years, is the fruit of more than one person. I cannot thank enough Emmanuel Boss, my advisor, for tremendous source of inspiration, unparalleled responsiveness, and support to put my work in perspective. I have been lucky to meet Emmanuel on the Navire Océanographique Tara between Cyprus and Malta, as besides having an advisor, I also had a friend. I have many memories of kayaking (paddle boarding for Emmanuel) the rivers of Maine with him, his enthusiasm, and tireless energy.

I would like to thank the other members of my committee for their advice: Lee Karp-Boss, Andrew Thomas, Jason Graff, and Fei Chai. Lee has been a continuous source of encouragement and provided key suggestions in structuring my writing. Andrew led me into the exploration of Maine by hiking Kathadin and with whom I had insightful discussions. Jason generously shared his cytometric data and tirelessly explained me it's deepest secrets, he was a great teammate on the NAAMES and EXPORTS campaigns. Fei Chai provided distinct ideas and advices.

I also wish to thank my officemate Alison Chase for teaming up at work and on the trails. Thank you to my fellow students and friends for their support: Robert Boenish, Guillaume Bourdin, Ashley Charleson, Matthew Hall, Mackenzie Mazur, Binbin Jiang, Javier Moreno, Mattie Rodrigue, Lino Sander de Carvacho, Drajad Seto, Constantin Scherelis, Jordan Snyder, Juliana Tavora, and Haley Viehman. Special thanks to James Loftin for assistance in setting up and making instruments. I would like to thank Shelby Helwig for proofreading my thesis and supporting me during the last several months of confined thesis writing. Many thanks are also due to the School of Marine Sciences office staff: William Ellis, Jodie Feero, Jessie Gunning, and Carrie Love. I would like to thank the entire NAAMES, PEACETIME, and Tara teams for great camaraderie and insightful

discussions and in particular Michael Behrenfeld. I acknowledge the crew of the R/V Atlantis, R/V Roger Revelle, and N/O Pourquoi Pas ? for their outstanding support at sea.

This work was funded by the National Atmospheric and Space Administration (NASA) under the NNX15AE67G and NNX15AC08G grants.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Over the past decades, optical measurements have transformed our perception of the ocean's biogeochemical processes. The capacity of optical sensors to sustainably observe key parameters like biomass, primary productivity, and diversity (*Bax et al.*, 2019) at a global scale with high temporal resolution is the primary reason optical measurements have made this impact. The measured optical properties (e.g. remote sensing reflectance, attenuation, backscattering) are used as proxies for many biogeochemical variables, for example, particulate organic carbon (*Rasse et al.*, 2017), phytoplankton biomass (e.g. chlorophyll a (*Davis et al.*, 1997), phytoplankton carbon (*Graff et al.*, 2015)), particle size (*Boss et al.*, 2001a; *Slade and Boss*, 2015), phytoplankton size (*Ciotti et al.*, 2002; *Houskeeper et al.*, 2020), or suspended sediments (*Dierssen et al.*, 2006a).
Transmissometers were the first optical instrument deployed on profiling platforms and underway systems of research vessels. They allowed researchers to observe intense horizontal and vertical gradient of bio-optical variables since the 1960s (*Dickey*, 1991). Later, the launch of satellites, exemplified by the Coastal Zone Color Scanner (CZCS, 1978), made estimating phytoplankton biomass and primary productivity of our oceans from space feasible (Behrenfeld and Falkowski 1997). More recently, autonomous platforms (e.g. gliders, floats, saildrones, *Chai et al.*, 2020) embedding bio-optical sensors were seeded throughout the ocean and provide a synoptic view beyond the surface ocean (e.g. biogeochemical-Argo network, *Claustre et al.*, 2020). Understanding how oceanic particles affect the inherent optical properties (IOP) of the ocean is fundamental to appreciate the variability in biogeochemical variables already derived from optical measurements and to derive new parameters.

The contribution of oceanic particles to IOPs has been extensively studied (e.g. *Durand and Olson*, 1996; *Green et al.*, 2003). While the underlying assumptions of optical models

varied between studies, most studies used Mie theory (*Mie*, 1908; *Bohren and Huffman*, 1983) to link particle size, refractive index, and concentration to IOPs. Mie theory models the optical properties of homogeneous spheres from their size and (complex) refractive index. These studies suggest that phytoplankton accounts for up to 60 % of the particulate beam attenuation while the remaining signal was attributed to non-algal particles (NAP). In the surface open ocean, where phytoplankton concentrations are relatively high compared to inorganic particles, these findings explained the correlation between particulate attenuation and phytoplankton biomass (*Behrenfeld and Boss*, 2006). On the other hand, these models typically underestimated the measured backscattering (*Stramski et al.*, 2004; *Dall'Olmo et al.*, 2009). Recent studies suggest that the missing backscattering resides in the structural complexity of particles and that the assumption of homogeneous sphere to represent marine particle is too simplistic (*Organelli et al.*, 2018). Despite these recent findings, the relations between phytoplankton, particulate attenuation, and backscattering are not well understood in the open ocean as datasets comprising particulate and phytoplankton size distribution with concurrent measurements of particulate attenuation and backscattering are lacking (*Jonasz and Fournier*, 2007). Backscattering is of much interest as, unlike beam attenuation, it is one of the products obtained from ocean color remote sensing. In this thesis, I build and assess such datasets for a broad range of environmental conditions (Chapter 2), to estimate the contribution of phytoplankton and NAP to particulate attenuation and backscattering measurements (Chapter 3) as well as the contribution of mesopelagic organisms to backscattering (Chapter 4).

The second chapter focuses on the size distribution of phytoplankton. As virtually every aspect of the life of phytoplankton is affected by its size, size is considered a master functional trait (*Chisholm*, 1992; *Litchman and Klausmeier*, 2008; *Weithoff and Beisner*, 2019). In fact, it's commonly used in trait-based approaches to advance our understanding of plankton ecology (*Litchman et al.*, 2007; *Weithoff and Beisner*, 2019). However, measuring phytoplankton size distribution is challenging as phytoplankton sizes spans

3.5 orders of magnitudes and phytoplankton cells must be distinguished from other particles of similar size (e.g. sediments, detritus). This makes the measurements of the complete size spectrum difficult as multiple instruments (e.g. cytometers) are required, need to be inter-calibrated, and then merged into a unified size distribution (*Laney and Sosik*, 2014; *Lombard et al.*, 2019). Here, I build a phytoplankton size distribution by calibrating and merging measurements from two cytometers: a classical one for small particles (< 10 µm) and an imaging cytometer for larger particles (> 7 µm). With this phytoplankton size distribution, I document the seasonal cycles and latitudinal patterns observed in the western North Atlantic and link them to environmental variables. In addition, I use the phytoplankton size distribution to assess current phytoplankton size indices derived from IOP and indirectly from ocean color satellites that have never been validated with independent size measurements due to the limited availability of such datasets (*Mouw et al.*, 2017).

Further, the third chapter uses the phytoplankton size distribution and together with concurrent observation of the full particulate size distribution quantifies the contribution of phytoplankton to the absorption, attenuation, and backscattering coefficients. This analysis relies on the computation of the efficiency with which cells and particles absorb and scatter light derived from models assuming homogeneous (Mie theory) or coated spheres.

Optical observations have been extensively used to retrieve parameters linked to phytoplankton but to a lesser extent to study other marine organisms. Diel vertical migration of mesopelagic organisms plays an important role in the biological pump by exporting particulate organic carbon from the euphotic layer of the ocean to the twilight zone (*Steinberg and Landry*, 2017). Traditional sampling required to study these organisms is largely conducted on board research vessels which have limited spatial and temporal coverage (e.g. research vessel rarely sample in the winter at high latitudes). Complementing such sampling with gliders (*Benoit-Bird et al.*, 2018; *Ohman et al.*, 2019) and profiling floats is of interest if they can provide information on the organisms'

distribution (*Boyd et al.*, 2019). In the fourth chapter, upon observing anomalies in particulate backscattering and ultra-violet fluorescence profiles typically associated with sinking aggregates (*Briggs et al.*, 2011), it became clear these anomalies could originate from another source. I determine that these anomalies are associated with the diel migration of mesopelagic organism. Diel and seasonal patterns across the Southern Ocean and North Atlantic, consistent with the behavior of zooplankton, emerged from the analysis of the biogeochemical float array.

The fifth chapter presents the development of a new tool to improve data acquisition from optical instruments. In the process of acquiring the necessary datasets for my thesis, I encountered difficulties with current software to log data from optical instruments (e.g. WET Labs ACS, ECO-BB3, and Sequoia LISST) as they were not intended for month long campaigns at sea. Manufacturer's solutions were designed for short-term deployment in laboratory or underwater deployments without data visualization. Novel approaches were needed to address the limitations in the manufacturer's solutions. Problems such as drifting timestamp, memory limitations, or the opposite, the generation of oversized files were limitations I had to overcome. Therefore, I added to the optical oceanographer's toolbox a logging software which was tested and improved during the multiple campaigns I participated in (e.g. NAAMES, PEACETIME, Tara Pacific, EXPORTS).

CHAPTER 2
DETERMINATION OF PHYTOPLANKTON SIZE DISTRIBUTIONS IN THE WESTERN NORTH ATLANTIC AND THEIR SEASONAL VARIABILITY

2.1 Introduction

Distinct planktonic communities have been associated with specific environmental conditions, suggesting that the ecological strategies and trait combination of these communities enable them to flourish (*Margalef*, 1978). These findings stimulated the application of trait-based approaches to advance our understanding in plankton ecology (*Litchman et al.*, 2007; *Weithoff and Beisner*, 2019). Phytoplankton traits that define their ecological niche have been grouped into four types: morphological (e.g. cell size, shape, coloniality), physiological (e.g. photosynthesis, nutrient uptake), behavioral (e.g. motility), and life history (e.g. resting stages) (*Litchman and Klausmeier*, 2008). As virtually every aspect in the life of phytoplankton is influenced by these traits, and at a different level by its size, cell size is considered a master functional trait (*Chisholm*, 1992; *Litchman and Klausmeier*, 2008; *Weithoff and Beisner*, 2019). Note that most traits, and in particular size, are involved in both bottom up processes (e.g. nutrient acquisition) as well as top-down processes (e.g. encounter-rate with predators).

The distribution of phytoplankton biomass across a wide range of cell size (> 9 orders of magnitude in volume) is a fundamental property of marine ecosystems. Phytoplankton size distribution (PSD) has ecological and biogeochemical implications. For example, in most communities where small cells dominate, the microbial loop dominates trophic pathways, while at times where larger phytoplankton become significantly more abundant a more "classic food chain" prevails in trophic interactions (i.e. direct energy transfer to mesozooplankton, *Pomeroy*, 1974; *Marañón*, 2015). Cell size also determines phytoplankton metabolism as nutrient uptake and light absorption are functions of cell size

(*Kleiber*, 1932; *Karp-Boss et al.*, 1996; *Brown et al.*, 2004). Therefore, the PSD characteristics relate size-diversity to ecosystem functioning.

Another critical aspect of measuring phytoplankton size distribution is to evaluate related ocean color remote-sensing products, which are the only platforms providing a synoptic view of ocean biology (e.g. scales of observation shorter than week and resolutions greater than 10 km, *Bax et al.*, 2019). Relating ocean color to size is based on a mix of empirical and theoretical models and their underlying assumptions (e.g. *Loisel et al.*, 2006; *Kostadinov et al.*, 2010). To evaluate and understand these relationships mechanistically and improve our current understanding of the contribution of phytoplankton to inherent optical properties (IOP) of the ocean (e.g. missing backscattering enigma, *Stramski et al.*, 2004) it is essential that phytoplankton specific optical properties are measured concurrently with their size distribution (*Jonasz and Fournier*, 2007).

Phytoplankton cell diameter spans 3.5 orders of magnitude from the smallest cyanobacteria (with an equivalent spherical diameter (ESD) of 0.5 µm) to the largest dinoflagellates ($ESD = 2000$ µm, *Noctiluca*). This is similar to the difference in length between an ant (5 mm) and a blue whale (25 m). Given the size range of phytoplankton there is no single technique that captures their full size spectrums but rather a combination of techniques and instruments is needed, each with its own size range limitation (*Lombard et al.*, 2019). Typically, each instrument is biased at its lower and upper detection range. The smallest particles are too small to be counted or sized effectively (e.g. signal below detection limits, too few pixels) while larger particles are scarce, and the volume analyzed is often too small to quantify them in a statistically significant manner. Another challenge in measuring phytoplankton size distribution resides in the ability to distinguish between phytoplankton, heterotrophic protists, and non-algal particles (NAP). While marine particle size distributions have been measured for decades (starting with Coulter Counters,[*Sheldon et al.*, 1972), phytoplankton size spectra are still scarce and often limited to a narrow size range. The principal reason being that to date only two types of instruments

(cytometers and quantitative imagers) can size phytoplankton semi-automatically while distinguishing it from other particles. Traditional microscopy, can also be used to determine the abundance and cell size of larger phytoplankton (> 5 µm) (*Huete-Ortega et al.*, 2011), but manual size measurements under a microscope is time consuming and hence hard to perform on live samples. Finally, many phytoplankton species (e.g. belonging to cyanobacteria and diatoms) form chains and colonies which can be 1-2 order of magnitudes larger than the any single cell within them, making the concept of size/function relation even more complicated.

Here I build phytoplankton size distributions by merging cytometry and imaging cytometry observations during four phases (winter transition, accumulation, climax, and decline) of the western North Atlantic annual cycle, as part of the North Atlantic Aerosol and Marine Ecosystems Study (NAAMES, *Behrenfeld et al.*, 2019). I then parametrize the variability of the size distribution (e.g. exponent of the PSD, mean phytoplankton diameter) and find co-varying environmental descriptors such as physical (e.g. mixed layer depth, temperature) and chemical (e.g. nutrients) parameters. Current optical size algorithms based on a variety of strategies (e.g. relative pigment absorption, spectral slope, anomalous dispersion) are evaluated. Our findings provide a PSD dataset that could be used, in addition to addressing ecological questions, for example, to improve our understanding of the contribution of phytoplankton to the inherent and apparent optical properties of the ocean (addressed in Chapter 3).

2.2 Materials and Methods

2.2.1 Data Collection

A continuous flow-through system was installed on the R/V Atlantis with water pumped into the vessel from ~ 5 m depth using a diaphragm pump (*Boss et al.*, 2019). The system was equipped with optical sensors (of interest here are the Sequoia LISST, WET Labs ACS, and WET Labs ECO-BB3) and the McLane Imaging FlowCytobot

(IFCB). Discrete samples were collected from the flow-through system regularly and analyzed within a few minutes by the BD Influx Cell Sorter flow cytometer (ICS). For each of these samples, to maximize the relevant volume of seawater sampled by the IFCB (to increase the number of cells in each size bin), all IFCB samples within an hour and half and in the same water mass as the ICS sample were used to build a phytoplankton size distribution merging both instruments. The samples were assumed to be in the same water mass if no density or chlorophyll fronts were crossed. Fronts were defined as changes in chlorophyll *a* concentration by a factor greater than two within a 30 min interval as the ship was underway. This procedure resulted in 97 ICS-IFCB matched samples encompassing a wide range of physical-chemical conditions across four provinces of the North Atlantic (Arctic, Temperate, Subtropical Gyral, and Gulf Stream/Sargasso Sea), in different seasons, with surface chlorophyll concentrations varying by more than a factor of 100, from 0.05 to 5.14 µg/L, and mixed layer depth varying from 5 to 255 m (Figure 2.1).

Inherent optical properties were measured while underway and used here to establish a closure of the PSD and evaluate optical size algorithms. The optical package was setup and calibrated following the recommendations of Boss et al. 2019 and the calibration independent technique of *Slade et al.* (2010). Near forward particulate scattering was measured with a Sequoia LISST (Type B, $\lambda = 670$ nm), hyperspectral particulate attenuation and absorption were measured by WET Labs ACSs and particulate backscattering was measured with a WET Labs ECO-BB3 (λ =470, 532, and 650 nm). In addition, a thermosalinograph (Sea-Bird, SBE45) continuously measured temperature and salinity. The particulate size distribution was computed from the LISST data as in *Boss et al.* (2018a). The size index γ_{c_p} and $\gamma_{b_{bp}}$ which relates to the particulate attenuation (c_p) and backscattering (b_{bp}) spectral slopes to the particulate size distribution slope were computed following *Boss et al.* (2001b). Phytoplankton absorption spectra ($a_\varphi(\lambda)$) were computed with the spectral decomposition method of *Chase et al.* (2013).

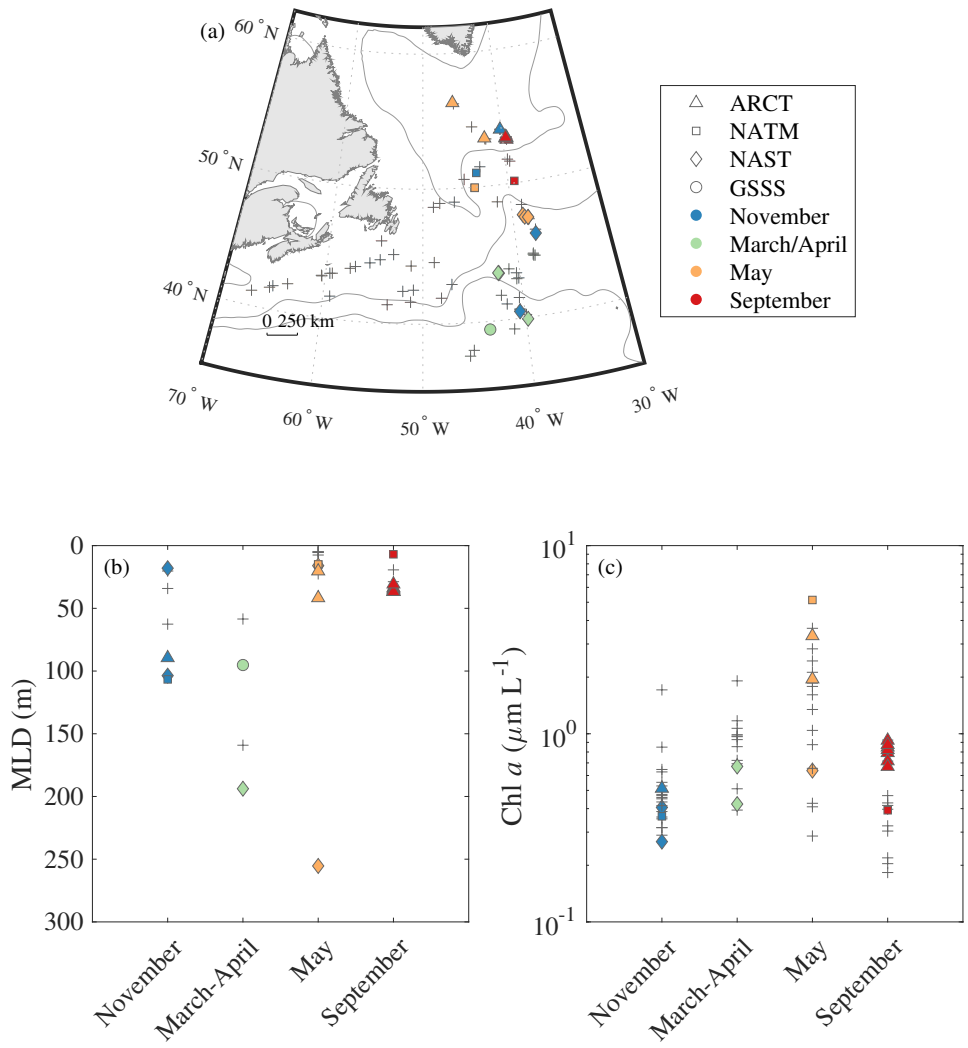

Figure 2.1. Environmental conditions of the NAAMES campaigns. (a) Location of the NAAMES cytometric samples in the western North Atlantic. Colored dots are used in the environmental analysis (Figure 2.17), while gray markers are PSD with limited contextual data. The subregions delimited by gray contour lines are from South to North: the Gulf Stream and Sargasso Sea province (GSSS), the North Atlantic Subtropical Gyral province (NAST), the North Atlantic Temperate province (NATM), and the Atlantic Arctic province (ARCT). Mixed layer depth (MLD) (b) measured at every station visited and chlorophyll *a* (c) measured at stations and underway are presented as function of seasons/NAAMES campaigns. Note that the deep MLD in May was observed in the core of an anticyclone (Station 4).

Contextual environmental data from the NAAMES campaigns used here were obtained as follows: the mixed layer depth (MLD) was computed for every CTD cast using the buoyancy frequency (N^2, Brunt-Väisälä frequency) according to Mojica and Gaube (in prep). The mean Brunt-Väisälä frequency ($\bar{N^2}$) between the surface and 100 m was also used as an indicator of the stratification of the upper water column. Inorganic nutrient concentrations (phosphate, nitrite+nitrate, and silicate) were measured from samples collected by the Niskin bottles of the CTD Rosette (*Baetge et al.*, 2020). Water samples for High-Performance Liquid Chromatography (HPLC) pigment analysis were obtained from samples taken from Niskin bottles and the underway flow-through system; the analysis was performed by the Ocean Ecology Laboratory at NASA following the protocols published by *Van Heukelem and Thomas* (2001) and (*Hooker et al.*, 2009). Stations occupied during the NAAMES campaigns were classified into four subregions based on the smoothed mean dynamic topography (MDT) defined by *Della Penna and Gaube* (2019) and are similar to the biogeochemical provinces defined in *Longhurst* (2007). The boundaries of the four subregions: the Gulf Stream and Sargasso Sea province (GSSS), the North Atlantic Subtropical Gyral province (NAST), the North Atlantic Temperate province (NATM), and the Atlantic Arctic province (ARCT) are shown in Figure 2.1. The daily photosynthetically available solar irradiance (PAR_{daily}) in the mixed layer ($< I_{ML} >$) was estimated with:

$$< I_{ML} >= \frac{PAR_{daily}}{K_{d_{PAR}} \times MLD} \left(1 - e^{-K_{d_{PAR}} \times MLD}\right) \tag{2.1}$$

$K_{d_{PAR}}$ the diffuse attenuation coefficient was computed from the chlorophyll *a* concentration with *Morel et al.* (2007). PAR_{daily} was modelled for clear sky conditions with the formula from *Frouin et al.* (1989).

2.2.2 IFCB Measurements of Abundance and Size Distribution of Large Phytoplankton (>7 µm)

The McLane Imaging FlowCytoBot (IFCB, serial number 107) uses laser-based cytometry coupled to a camera to capture images of particles as well as their side-scattering

(635 nm) and fluorescence (ex/em 635/680 nm) properties. Details on the fluidics and optics of the IFCB are given in *Sosik and Olson* (2007). Samples were pre-filtered with a 150 µm mesh to prevent blocking of the flow-cell, setting the upper size limit of cells sampled. Elongated particles, longer than 150 µm, such as Diatom chains were sampled but were rare (0.07 % of all particle counts). The lower size limit of cell sampled by the IFCB is defined by the resolution of the instruments (2.71 pixels/µm) and by the weak fluorescence signal of smaller cells (< 6 to < 10 µm, depending on instrument setting) which could go undetected by the IFCB. Therefore, the IFCB as set during the NAAMES campaigns imaged cells between 3 µm and 150 µm, covering two orders of magnitudes. To account for the decreased counting efficiency of the IFCB with smaller cells, I used an additional cytometer that is design for the enumeration of pico-phytoplankton (see section 2.2.3).

During the NAAMES campaigns the IFCB was operated in two modes: running discrete samples collected from Niskin bottles or continuously sampling from the underway system of the ship. The sampling volume was 5 mL and took ∼23 min to be analyzed. However, the volume analyzed varied between samples. This is because each time the instrument is triggered by a particle it takes some time to process it, time during which the IFCB cannot be triggered by another particle while the sample continues to flow through the chamber. To account for that, the actual volume analyzed is computed from the time during which the IFCB is available to be triggered. It ranged between 3 and 5 mL for the NAAMES samples and is a function of the concentration of particles in a given sample. The uncertainty of the volume analyzed is a combination of the accuracy of the 5 mL syringe of the instrument (±1 % according to manufacturer), the correct zero of that syringe (set at the beginning of each cruise with the instrument's software), and the precision of the instrument in measuring the processing time (unknown, hence not considered here). Because generally particle concentrations were low during NAAMES cruises (give ranges of mean particle concentrations for the cruises) several samples were

grouped together to increase statistical power. The average volume of matchup samples with the ICS was 29 mL (±9.7 standard deviations).

The IFCB was set to trigger only on fluorescent particles: mainly chlorophyll containing particles. For each image of a particle captured by the IFCB, the cross-sectional area is estimated by blob analysis of the images. The volume, surface area, and representative width of the particles are estimated using an updated version of the method described in *Moberg and Sosik* (2012) (code available at `https://github.com/hsosik/ifcb-analysis/tree/features_v3`, commit hash was 126658e412e50ba41a433f6805e41b0e423017fa). Other derived morphological features of each particle include: equivalent spherical diameter (ESD, based on cell volume assuming a sphere of the same volume), equivalent circular diameter (ECD, based on cell cross-sectional area, assuming a disk with the same cross section), major axis length, minor axis length, and Feret diameter (longest distance in between two parallel planes restricting the object perpendicular to the planes' direction). The retrieved features, in pixels, are then converted to µm using the instrument's pixel resolution. Pixel resolution was determined from size measurements of precision beads and was estimated to be 2.71 ± 0.04 pixels/µm (Figure 2.2.a). The size (in pixels) of the beads was determined after the blob extraction algorithm was applied to minimize effects (bias in shape and size) of such algorithm on the size calibration (*Giering et al.*, 2020). During each campaign, beads were run daily to monitor for potential drift in the resolution of the instrument and the resolution was stable across all four NAAMES campaigns (median absolute deviation = 0.1 µm, Figure 2.2.b). The uncertainty in ESD estimates was 0.1 µm which corresponds to the variability of the size of specific beads measured every day of the four campaigns. Note the resolution of the instrument (0.36 µm/pixel) was not included in the estimate of size uncertainty from the IFCB. I considered that the large number of particles sized (>3000 cells/sample on average) contributed to increase the instrument's resolution similarly to oversampling strategies applied in analog to digital converters (c.f. electronic literature).

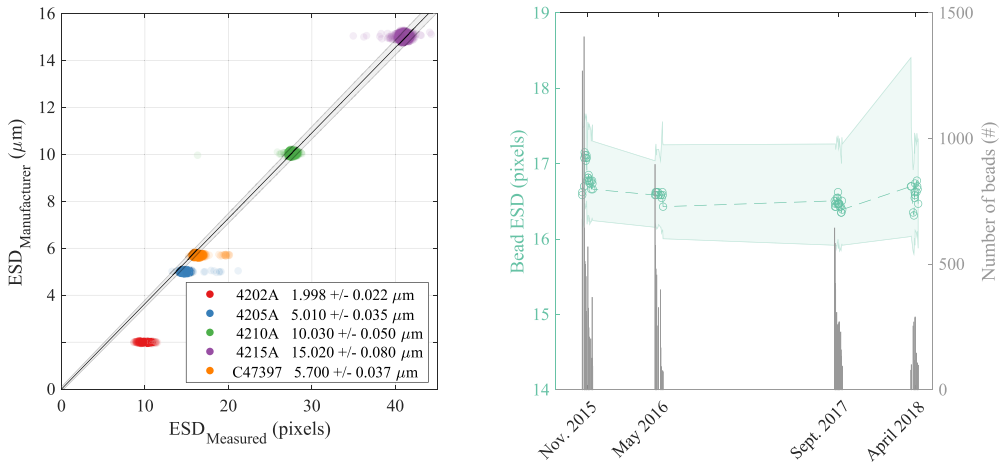

Figure 2.2. IFCB Calibration. (a) The equivalent spherical diameter (ESD) of beads specified by the manufacturer is plotted as a function of the diameter of beads measured by the IFCB. The beads used for the calibration were Duke Scientific 4202A (red), 4205A (blue), 4210A (green), 4215A (purple), and Life Technologies C47397 (orange) with an ESD of 1.998 ± 0.022, 5.010 ± 0.035, 10.03 ± 0.05, 15.02 ± 0.08, and 5.7 µm, respectively. To derive uncertainties noise was added to the manufacturer's size using a random number generator with a normal distribution set with the mean and the standard deviation specified by the bead manufacturer. The regression equation forced by zero is: $ESD_{\mu m} = 0.3638(\pm 0.0055) ESD_{pixels}$. (b) The ESD of the beads (blue dots) is plotted as a function of time to monitor the stability of the resolution of the instrument as a function of each expedition. The blue envelope represents the 5th and 95th percentile. The red histogram indicate the number of beads per sample run. The median absolute deviation (MAD) for all the NAAMES bead C47397 runs is 0.3 pixels corresponding to 0.1 µm.

For every sample, captured images were then classified manually using EcoTaxa, a web-based platform for image curation and annotation (*Picheral et al.*, 2017). While most images were living phytoplankton (86 %) because I triggered on fluorescence, other detected particles included zooplankton, non-living particles (e.g. detritus, plastic), or artefact (e.g. beads, bubbles) (Figure 2.3). Particles classified as artefact are excluded from the analysis and a sensitivity analysis on the PSD was conducted to observe the effect of excluding and including detritus and zooplankton on the PSD as those could not be separated from the ICS dataset. Including this type of particles biased the abundance PSD by 0.005 % on average and the volume PSD by 7 % on average.

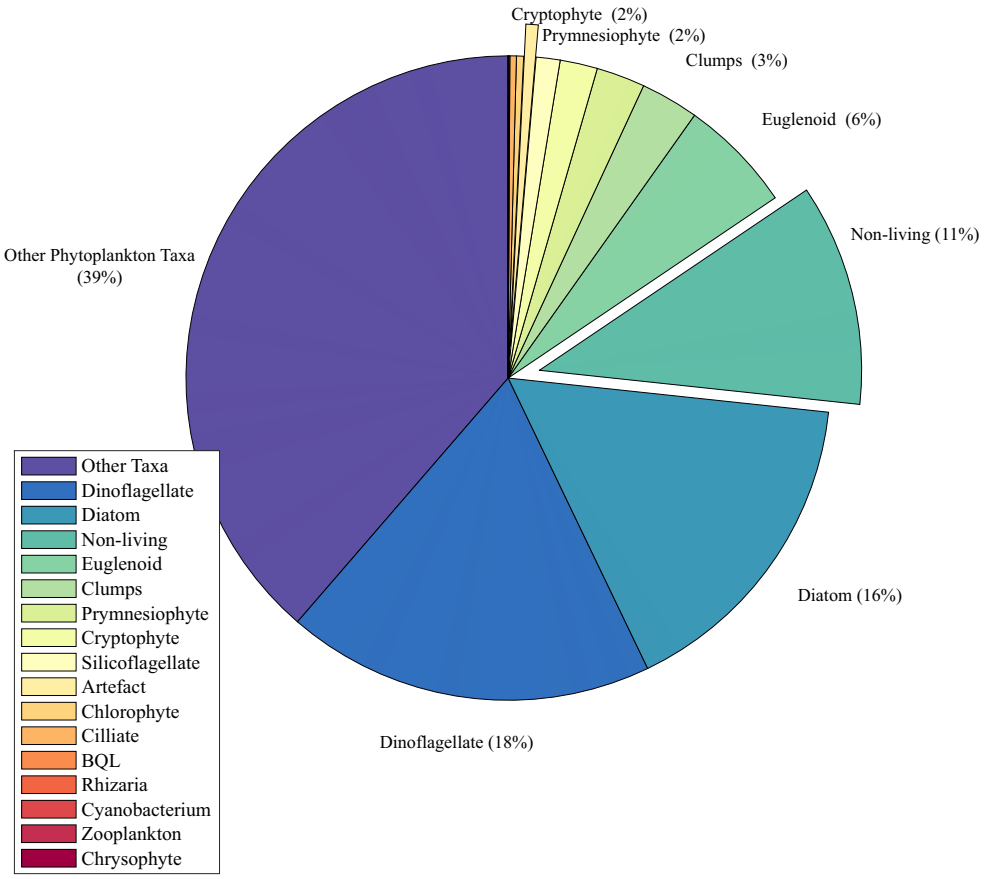

Figure 2.3. Taxonomic classification of particles greater than 10 µm imaged by the IFCB from samples matching the ICS during the four NAAMES campaigns. Other Taxa correspond phytoplankton taxon that could not be identified with manual classification. Particles below quantification limit (denoted as BQL), corresponds to images in a size range not well quantified by the instrument.

2.2.3 ICS Measurements of Abundance and Size Distribution of Small Phytoplankton (<10 µm)

The BD Influx Cell Sorter flow cytometer (ICS) detects particles in a flow stream passing through a nozzle of chosen size (100 µm) in front of a laser beam (488 nm). The ICS was set to trigger on particles fluorescing (Ex/Em = 488/692 nm) and recorded the peak and integrated values of forward scattering (FSC), side scattering (SSC), fluorescence at 530 nm (FL530), and fluorescence at 692 nm (FL692). From these parameters, cells were

classified into four major groups (*Prochlorococcus*, *Synechococcus*, pico-eukaryote, and nano-eukaryote) by gating them using the FlowJo Software. The configuration of the instrument can observe particles as small as 0.2 µm according to the manufacturer. Samples were pre-filtered through a 64 µm mesh setting the upper size limit of the ICS measurememts. On the lower end of the spectrum, the attenuation of the laser can be too low to be detected by the PMTs even when they are well tuned and, thus, it is possible to not observe small and high-light acclimated phytoplankton, specifically *Prochlorococcus*. Therefore, the operational lower size measured by the ICS is closer to 0.6 µm (*Lombard et al.*, 2019). Due to the low volume analyzed for which data was collected (0.8 ± 0.6 mL on average during the NAAMES campaigns) and the low abundances of bigger cells in oligotrophic waters the estimated concentration of cells in the upper end of the size spectrum has high counting uncertainties (c.f. Section 2.2.4.2). The volume sampled is calculated using the duration of the run and the flow rate. The flow rate was determined after each sample run by determining the loss of a 1 mL sample over a given time (typically greater than 60 seconds). The volume loss is determined using a pipettor, hence the accuracy in the volume sampled by the ICS is closely related to the accuracy of the pipettor used 1-2 µL. Based on the above the uncertainty in volume sampled by the ICS is 0.1 %. As there is no instrument downtime succeeding the trigger by a particle, the volume analyzed by the ICS is the same as the volume sampled.

The size of particles observed by the ICS has commonly been estimated using the time of flight (time during which the laser beam is forward scattered by the particle) or by measuring the FSC of the cell (*Gin et al.*, 1999; *DuRand et al.*, 2001; *Green et al.*, 2003; *Laney and Sosik*, 2014). It has been shown that the latter works well for cells that are approximately spherical while the retrieval of the size from elongated shapes, or when the index of refraction vary significantly within a sample is more difficult. The main reason being that FSC is a function of the index of refraction, shape, and sizes of particles; size having the most influence on the forward scattering (FSC) signal (*Jerlov*, 1976; *Gin et al.*,

1999). Due to instrument setup, the time of flight could not be used in this study, therefore the peak FSC by a particle is used to size cells with the ICS. However, one needs to realize that estimated sizes may be biased if cells have a significantly different index of refraction (e.g. *coccolithophores*) or shape such that they scatter differently than the phytoplankton in cultures used for size calibration (in this case *Dunalliela*, *Isochrysis galbana*, and *Thalassiosira weissglogii*).

The peak FSC is recorded by photomultipliers (PMT) which are used to amplify the dim signals from single cells of phytoplankton. The measured FSC signal in engineering units is function of the PMT gain settings and varies slightly from run to run (and even more in between deployments of the instrument). For these reasons FSC values of samples are normalized by FSC values of beads (used before most series of sample runs), as suggested by *Green et al.* (2003) and *Gin et al.* (1999). Because beads were not run at all gain settings the normalization was further constrained using a *Dunaliella* culture for which the size was consistent, evidenced with IFCB measurements of the same culture (Figure 2.4). To retrieve the ESD from the normalized FSC, a series of cultures sized with the IFCB was used for the larger end of the size spectrum, while the lower end of the size spectrum was calibrated assuming an average Synechococcus ESD (Figure 2.5). The series of cultures (*Dunaliella*, *Scrippsiella*, and *Thalassiosira weissflogii*) were ran within two hours on both instruments to minimize temporal variability of the cultures. The size of *Synechococcus* used for the calibration of the ICS is based on direct size measurements of *Synechococcus* from published literature (Table 2.1). The average value used was 1.17 µm (with 10^{th} and 90^{th} percentile of 0.83 and 1.48 µm respectively), the variability in the *Synechococcus* size could be due to the differences between ecotypes, cells physiology (light acclimation: ±0.1 µm *Morel et al.* (1993), nutrient availability: ±0.3 µm *Heldal et al.* (2003)), and cells diel division (±0.1 µm, Figure 2.6, in agreement with *Olson et al.* (1990)). The uncertainties in the retrieval of the ESD from the measurement of FSC by the ICS

(Figure 2.5) was estimated by propagating the uncertainty in the normalization of the FSC, the measured variability in FSC, and the variability in the size estimate of the cultures.

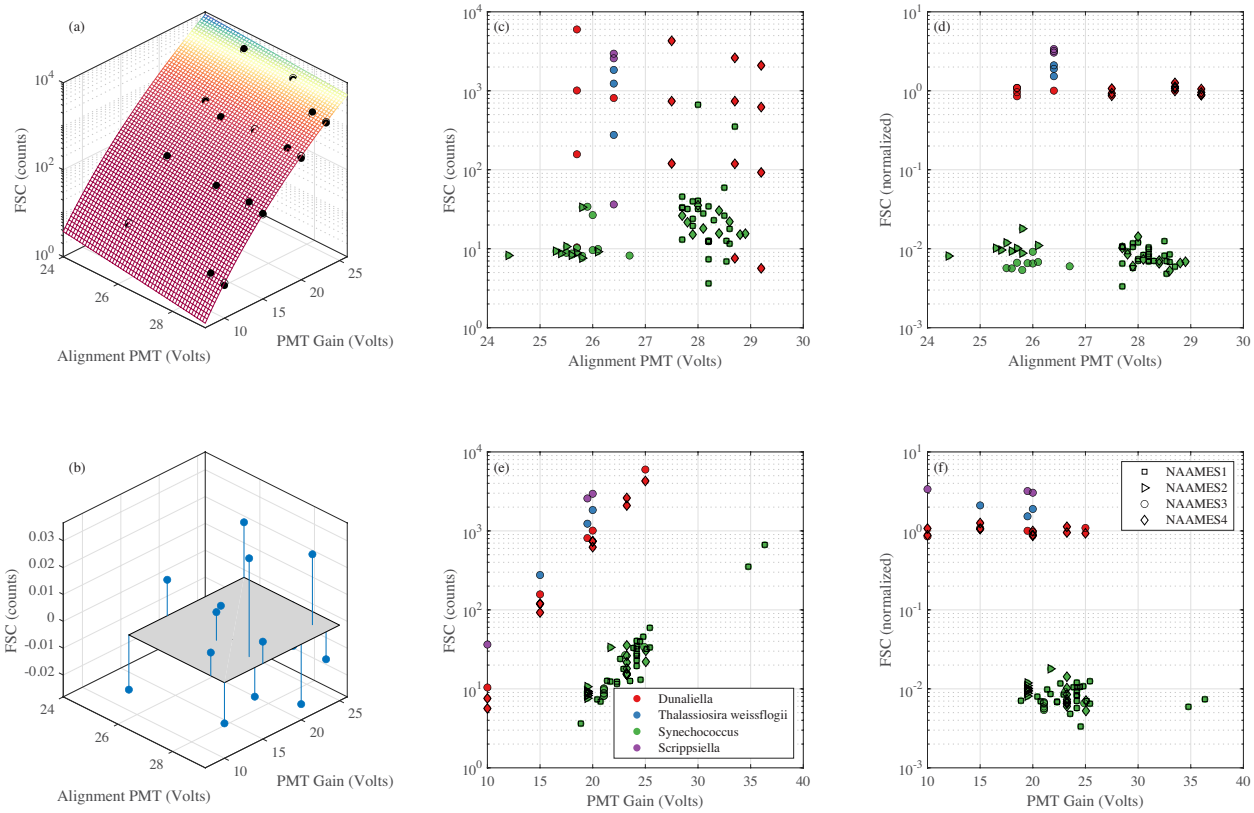

Figure 2.4. ICS forward scattering peak (FSC) normalization. (a) Forward scattering peak measured by the FCM for *Dunaliella* cultures (black dots) as function of PMT gain and alignment. The fitted surface is used to normalize the FSC and obtain an FSC independent of instrument settings. The residuals of the fit are presented in (b). (c) and (e) show FSC values for four cultures as a function of the FCM PMT alignment and gain settings respectively. (d) and (f) are similar plots after FSC is normalized for instruments settings (by the curve shown in (a)). The culture used were *Dunaliella* (red), *Thalassiosira weissflogii* (blue), and *Scrippsiella* (purple). *Synechococcus* (green) come from surface samples of the four NAAMES campaigns.

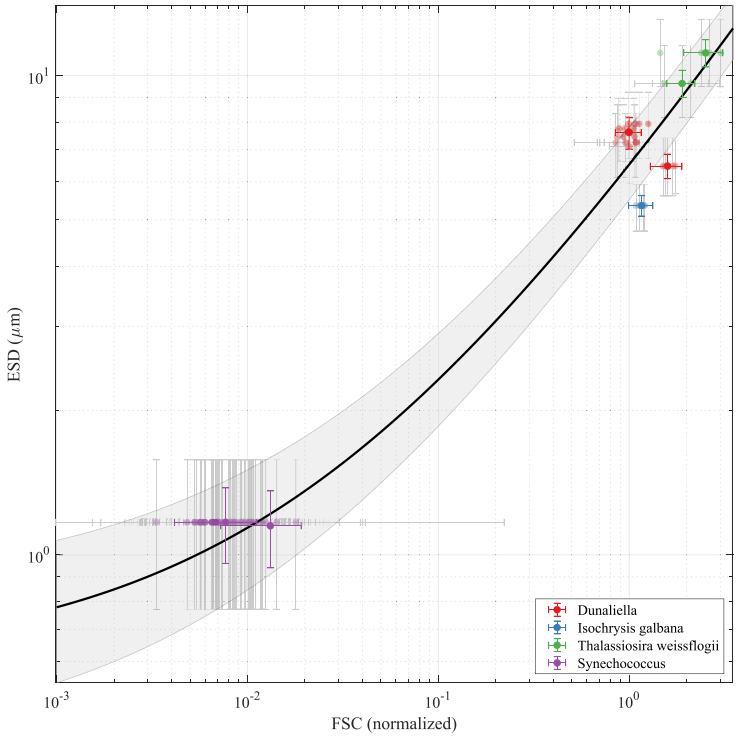

Figure 2.5. Calibration curve (black line) of the ICS normalized forward scattering peak (FSC) to size expressed in equivalent spherical diameter (ESD). The *Dunaliella* (red) and *Thalassiosira weissflogii* (blue) cultures were sized with the IFCB while the size of *Synechococcus* is based on a literature review (Table 2.1). The *Synechococcus* were gaited from the surface samples of the four NAAMES campaigns. The data is fitted to a first-order polynomial by a weighted linear least-squares type II (Matlab function lsqcubic from Peltzer, based on York 1966) after a Box Cox transformation (Box and Cox 1964). The equation of the fit is: $ESD_T = 2.06(\pm 0.15) \times FSC_T + 4.12(\pm 0.26)$ with Box Cox exponents $\lambda_{FSC} = 0.453$ and $\lambda_{ESD} = 0.754$. The uncertainty in the conversion from FSC counts to ESD are represented by the shaded area and are estimated to ±0.3 µm at 1 µm, ±0.5 µm at 2 µm, ±0.8 µm at 4 µm, and ±1.5 µm at 8 µm. Error bars correspond to the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile of the measured values on both x and y axis.

Table 2.1. Equivalent Spherical Diameter (ESD) of *Synechococcus* reported in past studies located in the North Atlantic region. Note that some studies reported other measurements (minor and major axis or volume) the computation of the ESD was carried assuming a prolate spheroid if needed. The median ESD is 1.17 ± 0.21 µm, the 10^{th} percentile is 0.83 µm, and the 90^{th} percentile is 1.48 µm. (*reported size in volume (µm^3))

Location	Strain	Type	Size Reported (µm)	ESD (µm)	Methodology	Reference
North East Coast of South America (8°44' N, 50°50' W)	SYN48 = CCMP833 = WH6501	natural	$0.9 - 1.3 \times 1.8 - 2.2$	1.34	epifluorescence microscopy	*Waterbury et al.* (1979)
Georges Bank and Sargasso Sea	similar SYN48	natural	0.8-2.0	1.4	epifluorescence microscopy	*Waterbury et al.* (1979)
Sargasso Sea, Narragansett Bay, R.I. shelf, Barbados Deep, Grenada Deep	Chroococcoid cyanobacteria, Type I	natural	0.8×1.0	0.86	transmission electron micrographs	*Johnson and Sieburth* (1979)
Sargasso Sea, Barbados Deep, Grenada Deep	Chroococcoid cyanobacteria, Type II	natural	0.7×0.9	0.76	transmission electron micrographs	*Johnson and Sieburth* (1979)
Sargasso Sea	Chroococcoid cyanobacteria, Type III	natural	0.4×0.5	0.43	transmission electron micrographs	*Johnson and Sieburth* (1979)
Woods Hole, East Coast, USA	WH8018 = L1604	culture	1.6	1.6	Coulter Counter ZBI	*Bricaud et al.* (1988)
Western North Atlantic, South Atlantic	WH8011, WH8112, WH8103, WH8113, WH8401, WH8406	culture	$0.7 - 0.9 \times 1.25 - 2.5$	1.06	phase-contrast photomicrograph, transmission electron micrograph	*Waterbury et al.* (1985)
Sargasso Sea (High irradiance)	MAX01	culture	0.812	0.81	Coulter Counter	*Morel et al.* (1993)
Sargasso Sea (Low irradiance)	MAX01	culture	0.924	0.92	Coulter Counter	*Morel et al.* (1993)
Sargasso Sea (Low irradiance)	MAX41	culture	0.853	0.85	Coulter Counter	*Morel et al.* (1993)
Bermuda Atlantic time-series (31°50' N, 64°10'W)		natural	0.87 (0.74-1.22)	0.87	flowcytometry	*DuRand et al.* (2001)
34°N 6°W	WH8012	culture	0.87	0.87	epifluorescence microscopy	*Bertilsson et al.* (2003)
Sargasso Sea	WH8103	culture	1.17	1.17	epifluorescence microscopy	*Bertilsson et al.* (2003)
Western North Atlantic 10 mmol L21 Mg11 ASW	WH7803	culture	1.6* +/- 0.2	1.45	electron microscope	*Heldal et al.* (2003)
Western North Atlantic 10 mmol L21 Mg11 ASW	WH8103	culture	1.02* +/- 0.07	1.25	electron microscope	*Heldal et al.* (2003)
Western North Atlantic (ASW)	WH7803	culture	$0.62 * \pm 0.07$	1.06	electron microscope	*Heldal et al.* (2003)
Western North Atlantic (ASW)	WH8103	culture	$0.83 * \pm 0.06$	1.17	electron microscope	*Heldal et al.* (2003)
Western North Atlantic (PCRS-11)	WH7803	culture	$0.71 * \pm 0.08$	1.11	electron microscope	*Heldal et al.* (2003)
Western North Atlantic (PCRS-11)	WH8103	culture	$1.2 * \pm 0.2$	1.32	electron microscope	*Heldal et al.* (2003)
North Atlantic (33.7423°N 67.4913°W)	WH7803	culture	1.8 volume	1.51		*Brown et al.* (2006)
North Atlantic (8°N 50°W)	CCMP833 = SYN48 = WH6501	culture	$0.8 - 1.06 \times 1.01 - 1.22$	0.99	microscopy	NCMA Bigelow Laboratory
North Atlantic (33.7423°N 67.4913°W)	CCMP1334 = WH7803 = NEPCC549 = RCC 752 = DC2	culture	0.968×2.407	1.31	microscopy	NCMA Bigelow Laboratory
North Atlantic (26.9667°N 83.4167°W)	CCMP1768 = MSL001	culture	1.131×1.451	1.23	microscopy	NCMA Bigelow Laboratory
North Atlantic (24.4°N 83.7°W)	CCMP1769 = MSL002	culture	1.051×1.561	1.2	microscopy	NCMA Bigelow Laboratory
North Atlantic (22.495°N 65.6°W)	CCMP2370 = WH8102 = RCC 539	culture	1.019×1.6	1.18	microscopy	NCMA Bigelow Laboratory
Puerto Rico (17.96°N 67.04°W)	CCMP2669 = BMB04	culture	1.515×2.363	1.76	microscopy	NCMA Bigelow Laboratory

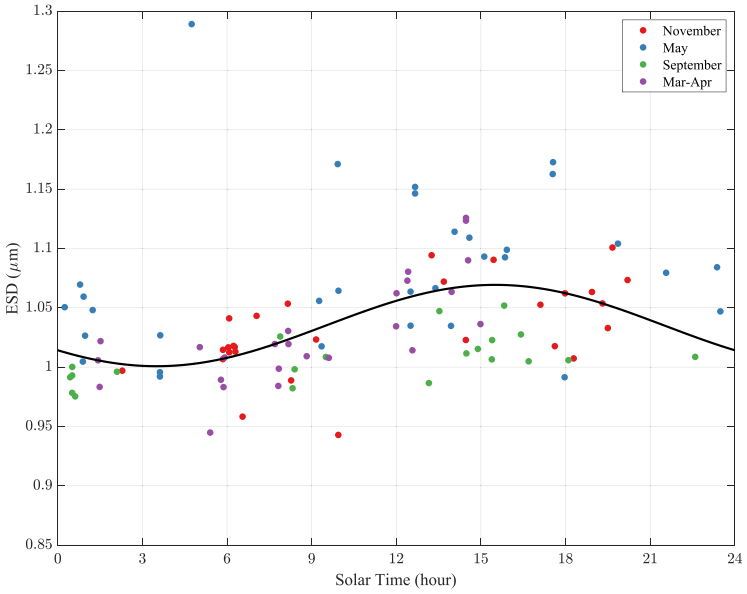

Figure 2.6. *Synechococcus* diel cycle. Median equivalent spherical diameter (ESD) of *Synechococcus* cells from surface waters as a function of the solar time (computed using the equation of time to convert UTC times) during the four NAAMES campaigns (I: red, II: blue, III: green, and IV: purple). The data is fitted to a sin function of the form $f(t) = a \times sin\left(\frac{2\pi}{24}t + b\right) + c$ by a non-linear least-squares method. The coefficients obtained are $a = -0.12$ µm; $b = 0.66$ unitless; $c = 1.19$ µm. Note that the dataset exhibit a minimum ESD of 1.07 µm at 3:28 solar time and a maximum ESD of 1.31 µm at 15:28 solar time, corresponding to a change in cross-sectional area and volume by a factor of 1.5 and 1.8, respectively.

Twenty three samples for which the FSC signal was saturated or below detection limit for a significant number of particles (detected manually from FSC vs. FL692 diagrams) and therefore biasing the size distribution, were excluded from the analysis. One additional sample was removed from the analysis as the ICS and IFCB disagreed on the measured cross-sectional area by more than four orders of magnitude in the overlapping size range of both instruments ($ESD \in [7, 10]$ µm).

2.2.4 Phytoplankton Size Distribution

2.2.4.1 Definition

The phytoplankton size distribution (PSD), also referred to as a size frequency or size spectrum, is defined as the number of phytoplankton cells per size interval per volume sampled (e.g., Jonasz and Fournier 2007):

$$n(D) = \frac{dN}{dD} \tag{2.2}$$

With dN the concentration of particles within the size interval $[D, D + dD]$. The unit of $n(D)$ is $\#\mathrm{mL}^{-1}\mu\mathrm{m}^{-1}$. The normalization by the bin width (dD) facilitated the quantification of variation in the PSD and comparison between instruments with different size bins. Depending on the application, one could derive PSDs for volume distribution ($\mu\mathrm{m}^3\mathrm{mL}^{-1}\mu\mathrm{m}^{-1}$, e.g. when studying biological processes) or cross-sectional area distribution ($\mu\mathrm{m}^2\mathrm{mL}^{-1}\mu\mathrm{m}^{-1}$, e.g. to model light absorption, scattering or attenuation through the water). They are defined as follow:

$$v(D) = \frac{dV}{dD}, \quad g(D) = \frac{dG}{dD} \tag{2.3}$$

With dV the volume concentration and dG the cross-sectional area concentration of particles within the size interval dD. These definitions of the PSD imply that a single parameter (D) can be used to describe the size of all phytoplankton regardless of their shape. Depending on the application and instrument used, a different definition of particle size may be used, and a synthetic list of the metrics relevant to marine particles is presented in *Jonasz and Fournier* (2007). Often, the diameter of a sphere with an equivalent volume to the particle, called an equivalent spherical diameter (ESD) has been used because it offers analytical simplicity by ignoring the particles shape and can easily be retrieved from widely used electro resistance techniques (e.g. Coulter Counter) that provide a measure of particle volume. In addition, most commercial instruments (e.g. Sequoia LISST) report size as the ESD, even if it's not directly measured (e.g. forward scattering technique is sensitive to the cross-sectional area of particles). The sensitivity to

different characteristics of the particles explains discrepancies in ESD measurements in between techniques especially for non-spherical particles (*Jennings and Parslow*, 1988; *Karp-Boss et al.*, 2007). In the context of ecosystem and biogeochemical functioning, studying phytoplankton size structure as a function of ESD is relevant. If interested in encounter-based processes (e.g. predator-prey interactions, nutrient fluxes) or particles behaviors in flows, other metrics such as the aspect ratio and the major axis length are more relevant (*Bergquist et al.*, 1985). This choice of metric is critical as it directly impacts the slope of the PSD and the classification of phytoplankton into size groups (c.f. Section 2.3.1). As mentioned above, the IFCB retrieves a variety of size metrics from images of cells allowing flexibility in the choice of size metric used. However, a proxy must be established for the ICS observations, the ESD was chosen as it's relevant to study biomass independently of the cell shape and for consistency with other studies. The volume and cross-sectional area are directly extracted from the images of phytoplankton for the IFCB, for the ICS they are computed assuming spherical particles:

$$V(D) = \frac{\pi}{6}D^3, \quad G(D) = \frac{\pi}{4}D^2 \tag{2.4}$$

2.2.4.2 Bin Size and Associated Uncertainty

The decision regarding the partitioning of the PSD into size intervals (bins), needs careful attention (*Jonasz and Fournier*, 2007). The highest resolution needed according to Shannon's Theory would be half the size of the smallest feature to be observed (*Shannon*, 1949). To determine the size of this feature - the bin size - for the present data set, I plotted the size distribution of each classified taxa and computed the associated standard deviation (σ) (Figure 2.7.a). All taxon tested had a similar σ of 1.3 ± 0.2 µm which was independent of their mean size (Figure 2.7.b). A direct consequence of choosing $\sigma/2$ as bin size is to exclude the ability to differentiate taxon with similar sizes or the size variability within a taxon (e.g. diel cycles). However, the bin size is most often limited by the uncertainties associated with the observation techniques (± 0.3 to ± 1.5 µm for the ICS and

± 0.1 µm for the IFCB) and the error associated with the number of particles observed within each bin due to the limited volume sampled. The latter was estimated using a continuous Poisson probability distribution function (*Jonasz and Fournier*, 2007) for different level of variability (σ) in the number of phytoplankton cells sampled (k) representative of the number of phytoplankton cells in the environment sampled (λ). I estimated the number of cells required to reach a given level of uncertainty (Figure 2.8).

$$P(k, \lambda) = \frac{\lambda^k \times e^{-\lambda}}{\Gamma(k+1)} \tag{2.5}$$

$$C(\lambda) = \int_{\lambda(1-\frac{\sigma}{2})}^{\lambda(1+\frac{\sigma}{2})} P(k, \lambda) dk \tag{2.6}$$

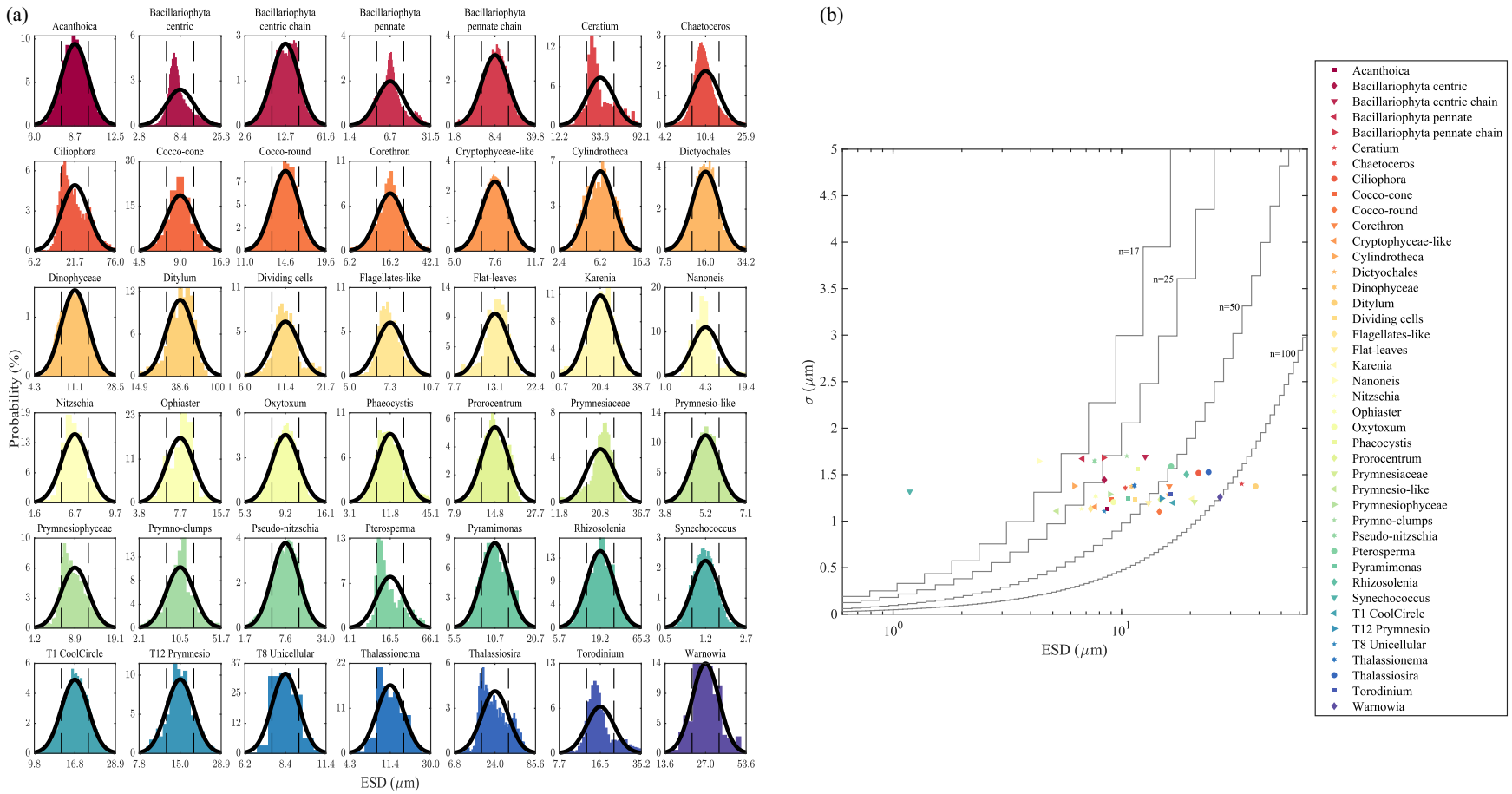

Figure 2.7. Size distributions of phytoplankton taxa encountered during the NAAMES campaigns (a). The x-axis correspond to the ESD (µm) and the y-axis is a probability. Each taxon presented here had more than 300 cells classified manually from the IFCB dataset and must have unimodal distribution. The *Synechococcus* sized with the ICS were added to that dataset. A normal distribution (black line) was fitted to the data (colored histogram) for each taxon. (b) Standard deviation (σ) of the normal size distribution fitted to each taxon is expressed as a function of the taxon's mean size (ESD). The bin width of the PSD is plotted as a stairstep line for different number of bins ($n \in [17, 25, 50, 100]$) as reference. Mean $\sigma = 1.3 \pm 0.2$ µm.

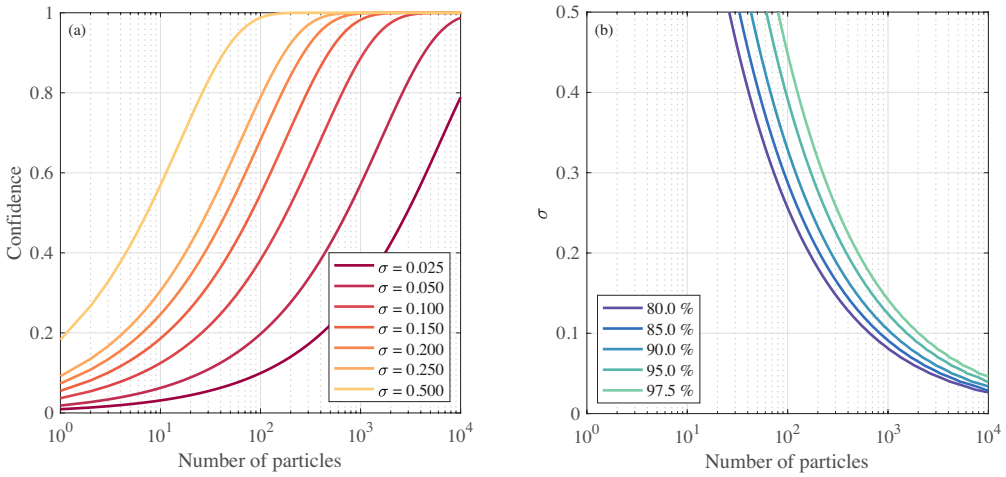

Figure 2.8. Cumulative continuous Poisson probability distribution within acceptable variability range applied to particles counted by an instrument (a). The curves are defined in equation 2.4 for different level of relative variability (parameter σ). (b) Relative variability in particles counted as a function of the number of particles counted expressed for different confidence levels.

Traditionally, size bins were distributed according to a power-law function. It presented the advantage of evenly distributing size bins on a log scale and made volume distribution nearly flat (*Sheldon et al.*, 1972). The downside of such partition of size bins is the low cell counts at the higher size bins resulting in high uncertainties for these bins (this is because if all bins have the same volume, the number of particles in each bin will be proportional to D^{-3}). Following this approach, I defined the edges of size bins (b_i) as follow:

$$q = \left(\frac{b_n}{b_0}\right)^{1/n} \tag{2.7}$$

$$b_i = b_0 q^i \quad i = 0, 1, .., n \tag{2.8}$$

With n the number of bins, b_0 and b_n the lower and upper boundaries of the size distribution. To reach the targeted grid resolution more than 100 bins would be required (Figure 2.7.b). This number of size bin requires an unrealistic volume to be sampled to maintain a reasonable level of uncertainty (< 50 %) given the cytometric instruments used (Figure 2.9.a-b). I then used traditional statistical binning algorithm to find an adequate number of bins that would reveal the shape of the underlying distribution. The average

number of bins for the NAAMES dataset was 17, 67, or 89 based on *Sturges* (1926), *Scott* (1979), and *Freedman and Diaconis* (1981) methods respectively (Figure 2.9.c). The uncertainty associated with these numbers of bin are still high (median σ for the last three bins is greater than 1.6) and more than 40 % of the bins have a width finer than the uncertainty in size. Therefore, I came to a compromise between a reasonable number of counts per size bins to minimize counting uncertainties, a bin width close to the instrument's resolution, and maximize the number of bins to achieve the desired resolution (Figure 2.9.b). The number of bins was set to $n = 12$ for the NAAMES dataset. This is lower than some other marine PSD (e.g. LISST PSDs are reported in 32 size bins).

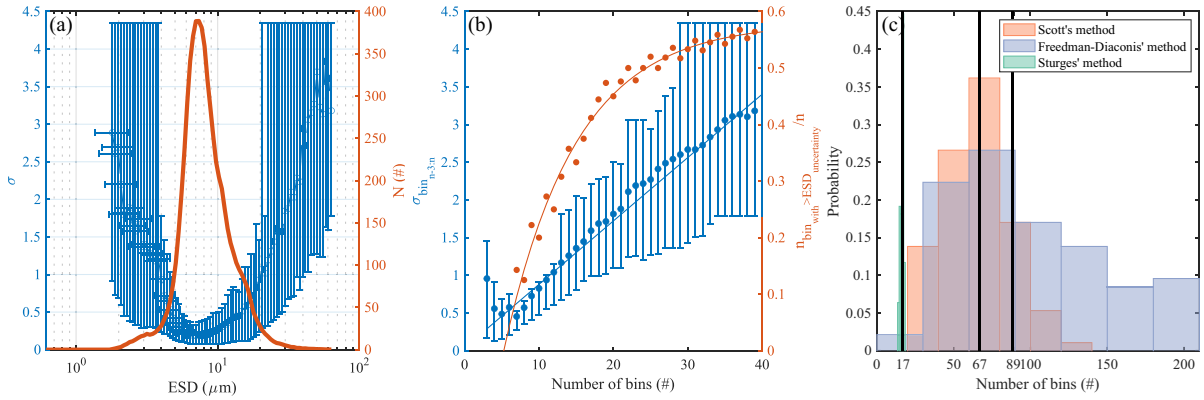

Figure 2.9. Counting uncertainties (σ, blue) estimated with Poisson probability distribution and number of particles counted per size bin (red) averaged for all PSD of the NAAMES dataset when setting the number of bins to 100 (a). Note that σ saturate to 4.3 when there is one cell per size bin and is not estimated when no cells are observed in a bin. The confidence used in the Poisson probability distribution is 80 %. (b) The average counting uncertainties of the last three bins of the PSD is plotted as a function of the number of bins used to make the PSD (blue dots). Error bars correspond to the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentile. (red dots) The relative number of bins with a width smaller than size uncertainty inherent to instruments used expressed as the number of bins. The lines were fitted to the data. (c) Histogram of the number of bin computed for each PSD using three binning algorithms designed to reveal the shape of the underlying distribution: (green) Sturges 1926, (red) Scott (1979), and (blue) Freedman and Davis (1981).

Size limits of the size distribution are traditionally set to the instruments' effective minimum and maximum observable size and depend on the size metric used. In our case $b_0 = 0.6$ µm and $b_n = 65$ µm for the size metric ESD, and $b_0 = 0.6$ µm and $b_n = 85$ µm for

the ECD (Table 2.2). The reported size for each bin is the geometric mean($\sqrt{b_i \times b_{i+1}}$). The smallest boundary (b_0) was set to the operational lower size measured by the ICS. The largest boundary (b_n) was set by the IFCB observations as the mean of the largest bin across all samples containing at least 10 particles. This choice of upper boundary led to excluding particles greater than the largest boundary of the grid of the PSD (109 particles over the entire dataset which account for 0.01 % of the total number of particles).

Center	Lower Edge	Upper Edge	Width
0.729	0.6	0.887	0.287
1.078	0.887	1.31	0.423
1.592	1.31	1.936	0.626
2.353	1.936	2.86	0.925
3.477	2.86	4.226	1.366
5.138	4.226	6.245	2.019
7.591	6.245	9.228	2.983
11.217	9.228	13.635	4.407
16.575	13.635	20.148	6.513
24.491	20.148	29.771	9.623
36.188	29.771	43.99	14.219
53.473	43.99	65	21.01

Table 2.2. Size of the bins for the phytoplankton size distribution (PSD) expressed in equivalent spherical diameter (ESD).

2.2.4.3 Merging PSD from Cytometers and Propagating Uncertainties

To merge the PSD from the ICS and the IFCB an overlap size range was defined between the highest "reliable" size from the ICS and the lowest "reliable" size from the IFCB. The upper end of the ICS spectrum is highly variable due to the low volume sampled and the low concentration of large cells (Figure 2.10.a-c). Therefore, the ICS upper size threshold were $b_h = [9.1, 10.8, 10.0, 7.4]$ µm for the size metric ESD during the NAAMES campaigns I, II, III, and IV, respectively. These thresholds were defined similarly to the PSD upper boundary for each campaign as significant differences between the campaigns were noted. The lower end of the IFCB size spectrum was defined as the median size maximizing the abundance PSD_{IFCB} for each campaign, as the IFCB miss

some of the smaller cells due to its PMT sensitivity (Figure 2.10.a-c). The lower size threshold of the IFCB were $b_l = [6.6, 7.1, 8.2, 6.5]$ µm, for the size metric ESD during the NAAMES campaigns I, II, III, and IV, respectively. Instrument settings (ICS and IFCB), differences in cell physiology (e.g. low chlorophyll/cell during NAAMES III) and variation in species composition between campaigns justify the choice of the ICS upper boundary and IFCB lower boundary on a per campaign basis. A weighted mean of the ICS and IFCB was applied in the merging area $b \in [b_l b_h]$ for $N(D)$, $G(D)$, and $V(D)$.

$$\begin{cases} n(b) = w_{ICS} n_{ICS}(b) + w_{IFCB} n_{IFCB}(b) \\ w_{ICS} = \frac{b_h - b}{b_h - b_l} \quad w_{IFCB} = \frac{b - b_l}{b_h - b_l} \end{cases} \tag{2.9}$$

With b_l and b_h being two bin sizes smaller than the lower boundary of the IFCB and two bin sizes larger than the ICS boundary, respectively. The uncertainties from each instrument were weighted following the same process. In two samples of the NAAMES I campaign the largest bin observed by the ICS was lower than b_l. In these cases, a robust linear regression was fitted to the two last bins of the ICS and the two first bins of the IFCB greater than b_l and used to interpolate the missing data.

Uncertainties in the PSD were estimated using a Monte Carlo approach that propagates uncertainties from multiple sources: counting uncertainty (Poisson's probability distribution function, PDF), size calibration uncertainty from both the ICS and IFCB, and uncertainties in the volume sampled by both cytometers. All Monte Carlo input parameters (Table 2.3) were varied independently assuming a normal distribution of width defined by the standard deviation of each uncertainty and assuming uncertainties are not correlated. The Monte-Carlo analysis was run for each PSD with 10^4 realizations. For each realization, the total average cross-sectional area ($\Sigma\bar{G}$), total biovolume (ΣV), the mean diameter based on number of particles, average cross-sectional area, volume ($D50_N$, $D50_{\bar{G}}$, $D50_V$), and PSD slope (m_φ) were computed. Doubling the number of realizations did not impact significantly any of these outputs. The 5^{th} and 95^{th} percentile of all realizations for each size bins were kept as lower and upper uncertainties for the PSD (Figure 2.10).

Table 2.3. Parameters and associated uncertainties used in the Monte-Carlo analysis of the phytoplankton size distribution uncertainties. (*Box-Cox transformed data.)

Parameter	Mean	Uncertainty	Reference
ICS calibration slope	1.8918 $tr_{fsc}/t_{\mu m}$*	0.2341 $tr_{fsc}/t_{\mu m}$*	Figure 2.5
ICS calibration intercept	3.2169 $t_{\mu m}$*	0.6551 $t_{\mu m}$*	Figure 2.5
IFCB resolution	2.7488 µm/pixel	0.1 µm/pixel	Figure 2.2
ICS volume sampled	Sample dependent	0.10 %	Estimated
IFCB volume sampled	Sample dependent	1 %	Manufacturer
N(D)	–	Poisson's PDF	Figure 2.8

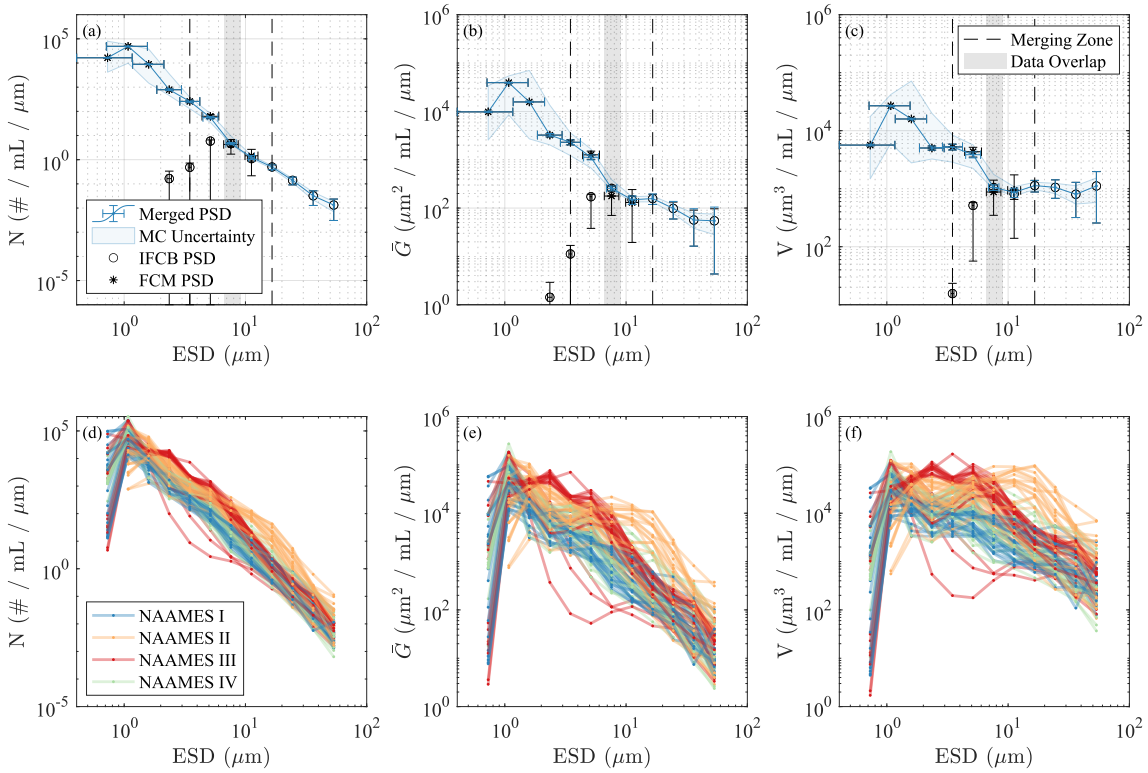

Figure 2.10. Phytoplankton size distribution (PSD) normalized by the bin width expressed in abundance (a), average cross-sectional area (b), and biovolume (c) as a function of the equivalent spherical diameter (ESD), for the 25$^{\text{th}}$ sample of the first NAAMES campaign (November 2015). The PSD from the ICS (blue asterisk) and IFCB (blue circles) are merged into one PSD (black line). The uncertainties on the x axis correspond to the instrument uncertainty in retrieving size (ESD). The error bars on the y axis represent the variability associated with counting statistics and the volume sampled uncertainties. The blue area correspond to the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile of the uncertainty propagation done with a Monte-Carlo. The vertical grayed area correspond to the overlapping section between the two instruments that is defined as "reliable". The dashed vertical line correspond to the merging area where both instruments are weighted averaged. (d-f) Merged PSD from the four NAAMES campaigns.

2.2.4.4 Particle Orientation in Cytometry

The orientation of particles in cytometers could potentially bias the average cross-sectional area PSD with respect to other observations of the environment. This cross-sectional area PSD is typically used in combination with optical measurements to understand the optical properties of the ocean. Optical instruments are either in-situ (e.g. instruments mounted on CTD Rosette, underway sampling) or remote (e.g. satellites). While in-situ instruments may be disturbing the ocean and certainly creating strong turbulences, the remote instruments observe the surface layer of the ocean as it is. The orientation of particles in marine environments was long assumed to be random (especially in optical modelling). However, under certain conditions when shear was low and turbulence was weak, phytoplankton were observed oriented near horizontally (*Malkiel et al.*, 1999; *Talapatra et al.*, 2013; *Nayak et al.*, 2018) and below the pycnocline diatom chains were progressively orienting vertically (*Font-Muñoz et al.*, 2019). Therefore, the cross-sectional area PSD needs to be computed for: (1) particles randomly oriented in the case of invasive instruments or strongly turbulent environment and (2) particles aligned with the flow for surface remote sensing in a weakly turbulent environment.

One could conceive that the design of a cytometer's chamber (e.g. sheath-flow principle) could result in higher probability to observe particles oriented with the flow. As particles rotate in a shear flow, the cytometer's chamber (e.g. sheath-flow principle) in our case, the time during which a particle is aligned with the flow increases with the aspect ratio of the particle (*Karp-Boss and Jumars*, 1998; *Karp-Boss et al.*, 2000; *Marcos et al.*, 2011). Hence, the probability to image particles oriented with the flow are higher and would then bias the observations of their average cross-sectional area. Taking advantage of the imaging capacity of the IFCB, the orientation of all the particles imaged during the NAAMES campaigns were analyzed (Figure 2.11). The orientation of a cell is defined here as the angle between the direction of the flow and the major axis of the ellipse that has the same second-moments as the blob. Small cells (< 7 µm) were evenly oriented in every

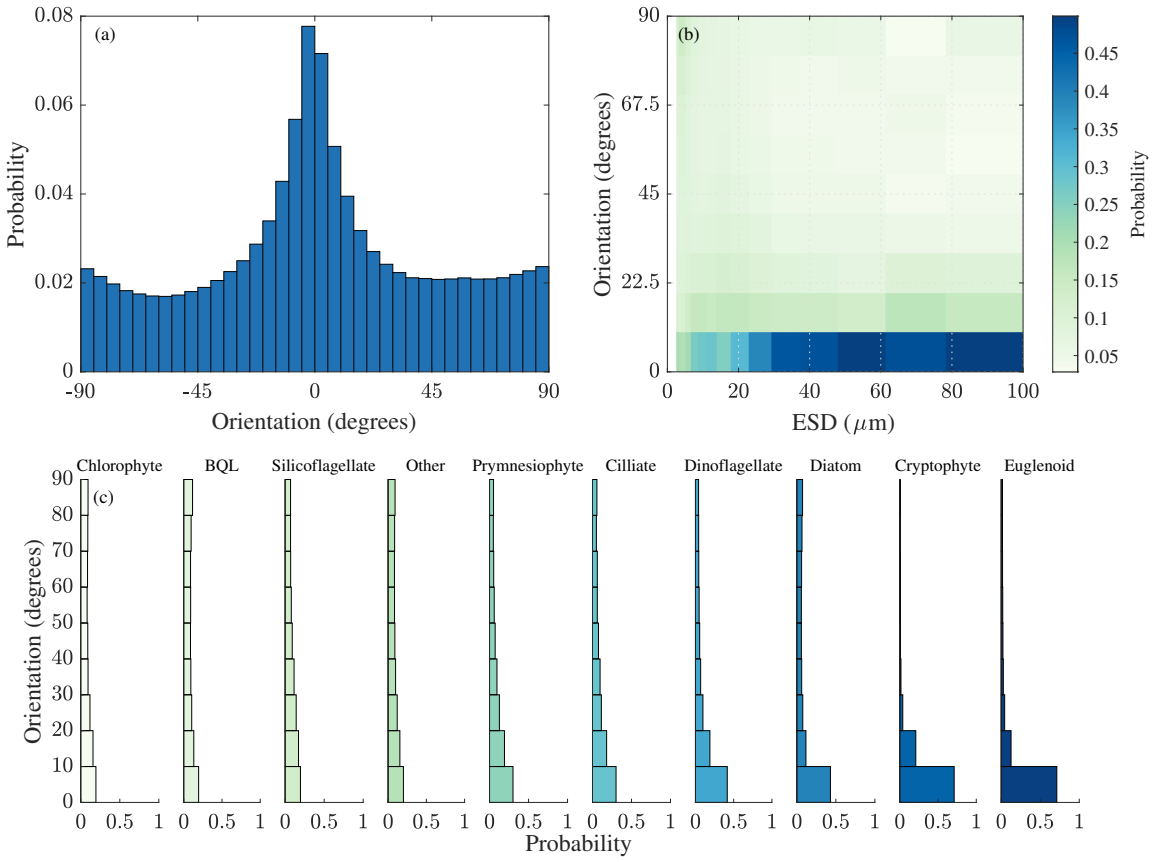

Figure 2.11. Particles orientation in the chamber of the IFCB during the four NAAMES campaigns ($n = 4.7 \times 10^6$) (a). Zero degrees correspond to the major axis of the particle being aligned with the flow. (b) Absolute orientation of the cells as a function of their size. All bins of a same size are normalized by the number of particles within that size range. (c) Probability of classified particles to be aligned with the flow. The group BQL (below quantification limit) corresponds to all cells with an $ESD < 8.8$ µm. The group Other correspond to cells that could not be classified due to the image quality. Groups with less than 1000 images are excluded from the plot (c).

direction, 28 % of the larger nano-cells (7-20 µm) were oriented with the flow, and 50 % of the larger cells were oriented with the flow (Figure 2.11.b). These results could be biased by the lack of resolution to adequately estimate the orientation of small cells. Nevertheless, these observations suggest that IFCB cell measurements, especially the cross-sectional area must be looked at carefully as they may not be representative of the average cell cross-sectional area but rather maximize it as the cells tend to be aligned with the flow.

In addition to the cross-sectional area of particles aligned with the flow (G), the cross-sectional area of particles randomly oriented ($\bar{G}$) was estimated from the surface area of the cells following a theorem of Cauchy that states: the mean projected area of a randomly oriented convex particle is one quarter of its surface area (*Cauchy*, 1832; *Vickers and Brown*, 2001). This method of retrieving the average cross-sectional area of phytoplankton was consistent with the work from *Jonasz* (1987), who estimated empirically the average cross-sectional diameter from the volume of marine particles (Figure 2.12.a). The cross-sectional area estimated from the surface area with Cauchy's theorem was compared to the directly measured cross-sectional area and no consistent offset was found as a function of particle size however a constant bias was found for specific taxa (Figure 2.12.b-c).

As the cross-sectional area from particles analyzed by the ICS is based on their ESD, no adjustments were needed.

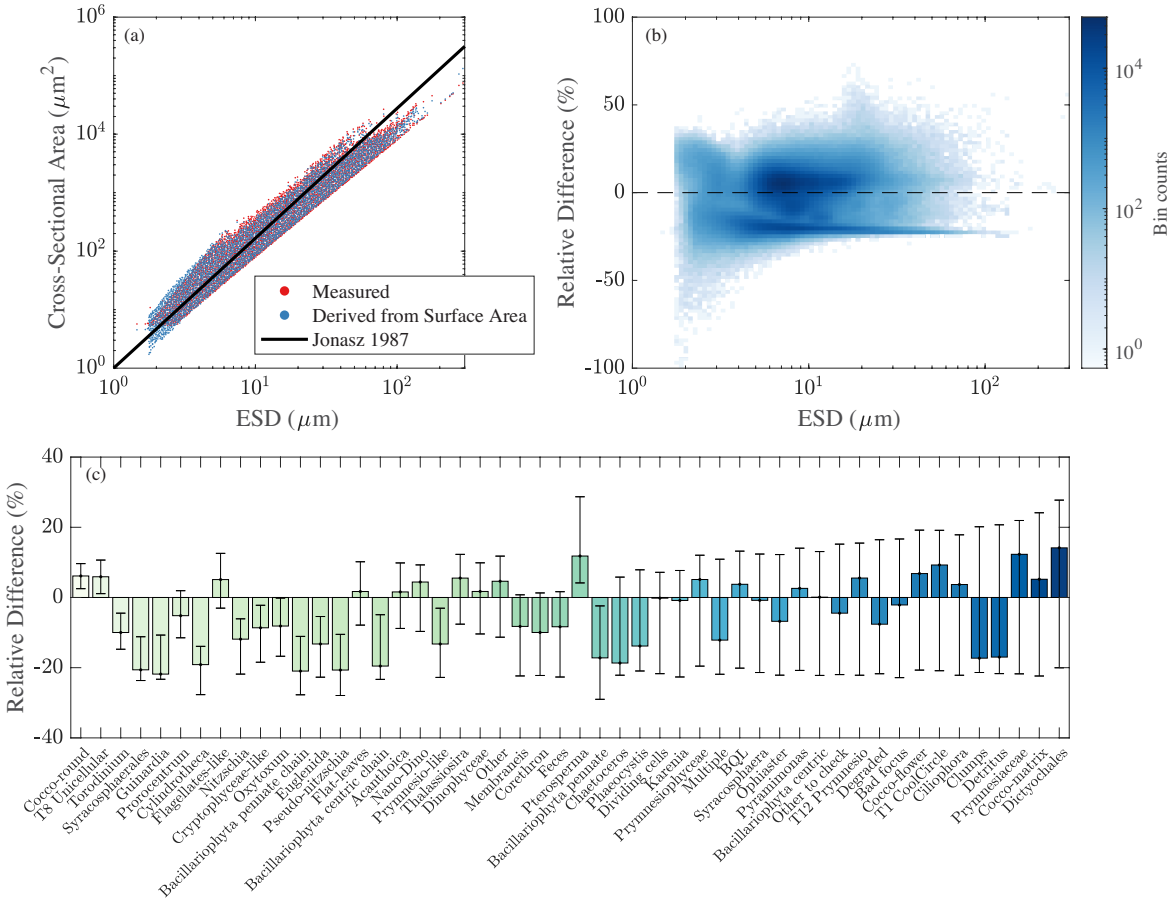

Figure 2.12. Comparison of the methods to retrieve the average cross-sectional area of marine particles observed by the IFCB during the four NAAMES campaigns ($n = 4.7 \times 10^6$) (a). (green dots) measured area by blob analysis of the image, (red dots) average cross-sectional area derived from the Surface Area applying Cauchy's Theorem, (black line) empirical model from Jonasz 1987 based on marine particles from coastal waters off Nova Scotia. Note that most green dots are masked by the red dots. (b) Relative difference between the average cross-sectional area derived from the surface area and the measured cross-sectional area as a function of the ESD. (c) Median relative difference between the measured cross-sectional area and the average cross-sectional area derived from the surface area for each taxon with more than 1000 cells validated. The error bar correspond to the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile. The median absolute deviation between the two methods is 3.3 µm^2 and no significant bias were observed.

2.2.4.5 Challenges of Measuring Size Distribution

While I report the major PSD uncertainties (counting, sampling volume, and size) and provide correction for cell orientation, other aspects linked to the measurements made to build the PSD need to be considered.

When sampling the environment, we alter it. In this study, I either sampled through the flow-through system of the ship using a low-shear pump or with Niskin bottles. The turbulence and shear field through which the particles were passing through was likely higher than the ambiant turbulence experienced by a phytoplankter in the ocean. This higher turbulence and shear field could break large particles (e.g. diatom chains) into smaller ones (*Gibbs*, 1981, 1982a,b; *Gibbs and Konwar*, 1983). This would directly impact the PSD by shifting it toward smaller cells and hence steepening the power-law-fit slope. However, the change in total cross-sectional area and biovolume are expected to be negligible. While I could not evaluate the effect of our sampling methods on the PSD, I compared the average size of phytoplankton and the total biovolume of phytoplankton between surface samples collected with the Niskin bottles and the flow-through system (Figure 2.13). Since both methods involve different type of handling and turbulence, if these had a significant effect, we would expect differences between them. No significant bias was found for samples that did not follow any treatment (samples used in the present study), similarly to what was observed by *Cetinić et al.* (2016) with diaphragm pump systems (I also used a diaphragm pump on all NAAMES cruises). Samples that were concentrated by filtering and re-suspending the particles underestimated the total bio-volume and a larger average cell size was observed. Therefore, any concentrated sample was excluded from the analysis.

Other considerations that might affect the PSD include sample aging (time between sample collection and analysis), non-homogenous particle distribution in the sample, high-shear flow of cytometric instrument used. The aging of my samples should be minimal in the order of minutes for samples from the flow through samples and less than two hours

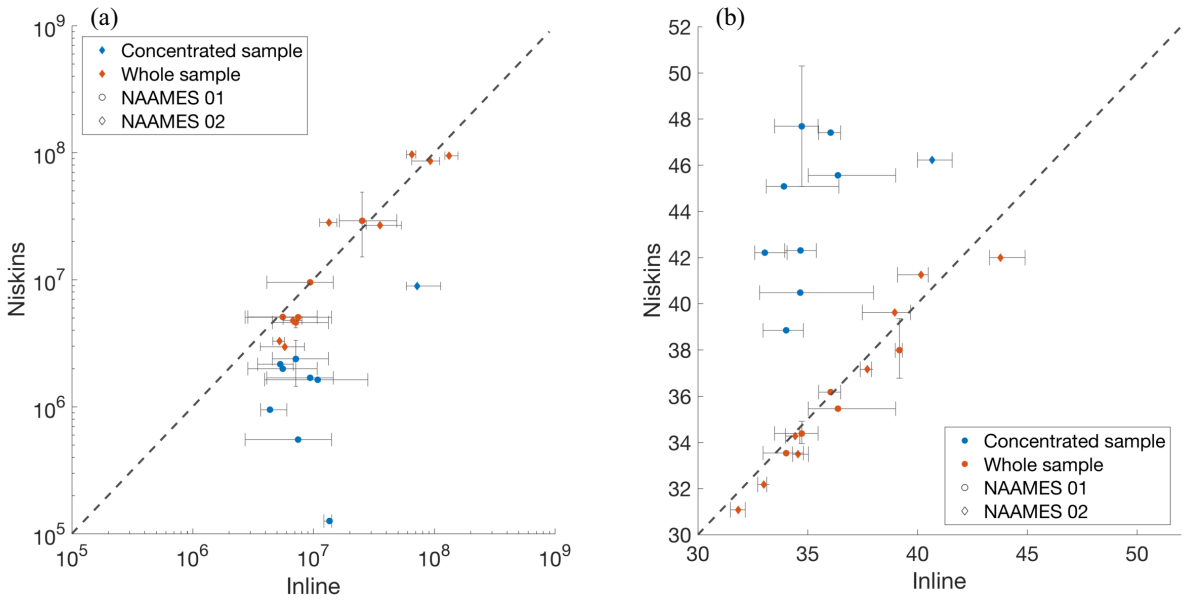

Figure 2.13. Comparison between underway samples and discrete samples from Niskin bottles. The total biovolume (a) and median equivalent spherical diameter (b) are compared. Niskin whole samples (red) were run directly after collection while niskin concentrated samples (blue) were concentrated using a filter. Error bars correspond to the 5th and 95th percentile of replicates. The dashed black line correspond to the 1:1 ratio.

for discrete samples from the Niskin bottles. The non-homogeneity of the water sampled or within the samples is discussed in *Jonasz and Fournier* (2007) Sections 5.6.3 and 5.6.4. Similarly, to the sampling method, the cytometers can break large particles due to the high shear flow caused by their fluidics and when the samples are pre-filtered through the 150 µm mesh at the intake. The impact of the cytometers on the particles observed was not evaluated here but is expected to be minimal. As most particles are not chains, they are not likely to break. Additionally, I still observed a substantial number of chains and a limited number of particle fragments with the IFCB. Note that in some instances the chains (e.g. *Chaetoceros*) observed by the IFCB did not entirely fit in the field of view of the camera resulting in an underestimation of their actual size.

2.2.4.6 Approximations

To simplify the inter-comparison of PSD and their integration in models, mathematical approximations of the PSD have been developed. These approximations are general descriptor of the size structure that indicate the contribution of each size classes to the total biomass. Commonly used approximations include a power law, a sum of lognormal functions, an average PSD shape, characteristic vectors, a Weibull function, an exponential function, and a sum of Gaussian and hyperbolic functions (*Jonasz*, 1983). The power law approximation has been routinely applied in studies on marine particles and is use here as well. The exponent of the power law is often referred as the Junge slope when the slope is typically close to -4 (*Junge*, 1963). A power law (Eq. 2.10) was fitted to the PSD by transforming the original data into log space, transforming equation 2.10 into a simple linear equation (Eq. 2.11).

$$n(D) = kD^m \tag{2.10}$$

$$\log n(D) = \log k + m \log D \tag{2.11}$$

Equation 2.11 was fitted to the data using a weighted robust type II least squares bisector regression (Matlab weighted robust fit function combined with MBARI lsqbisec function). The weights used for the fitting were inversely proportional to the PSD squared uncertainties on both dimensions. The fitting process is sensitive to the distribution of the size bins (Figure 2.14): with low number of size bins (< 8) the center of the bin used is not representative of the size of the particles within the bin resulting in a flattened PSD; as the number of bins increased above eight the mean PSD slope stabilized.

Another metric that I used to simplify the PSD to one parameter was the average diameter ($D50$) expressed for $N(D)$, $\bar{G}(D)$, and $V(D)$ giving $D50_N$, $D50_{\bar{G}}$, and $D50_V$ respectively. It corresponds to the diameter at which 50 % of the cumulative PSD is reached and therefore is smaller for larger $|m|$.

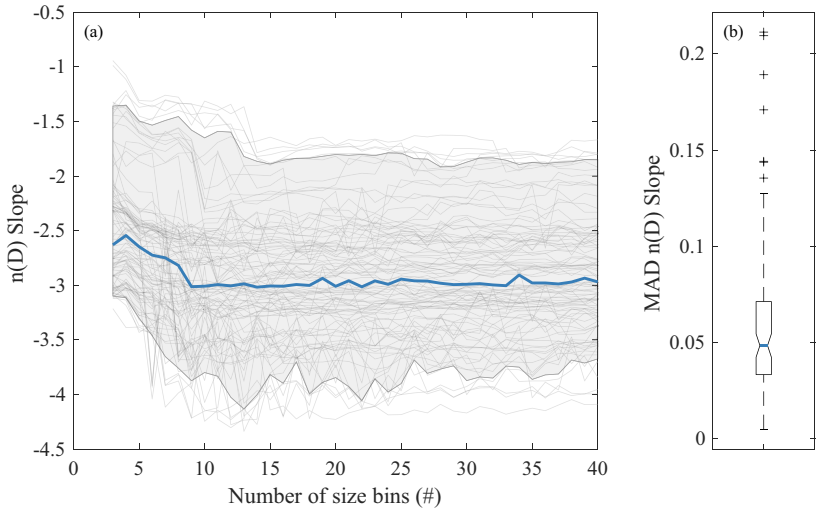

Figure 2.14. Sensitivity of the power-law-fit exponent (m) to the bin distribution of the PSD. (a) The slope of all the PSD are shown as light gray lines as a function of the number of bins ($n = 97$). The blue line correspond to the median slope of the dataset. The lower and upper bounds of the gray area correspond to the 5^{th} and 95^{th} percentile of the slope for the dataset. (b) The box plot of the median absolute deviation (MAD) of the slopes across all bin distribution tested (a). The central line of the box plot indicate the median, the edges of the box correspond to the 25^{th} and 75^{th} percentiles, and the whiskers extend to the most extreme data points excluding outliers (+). The slope median MAD for size bin distributions ranging from 3 to 40 bins is 0.05 ± 0.02.

2.2.5 Computation of Size Indexes from Inherent Optical Properties Based on Published Algorithms

From the concurrent measurements of inherent optical properties, I derived phytoplankton size indexes based on five published approaches: the biomass, the relative pigment concentrations, the absorption band effect, the spectral slope, and the volume scattering function.

The simplest method relate phytoplankton biomass to its mean size as they are known to co-vary. Hence, *Hirata et al.* (2008) (HIRATA08) proposed an empirical relation between chlorophyll a and a phytoplankton size index (S).

The co-variation of pigment packaging and concentration of accessory pigments with the size of dominant phytoplankton communities was first considered by *Ciotti et al.*

(2002), *Bricaud* (2004), and *Ciotti and Bricaud* (2006) (CB06). In this method, a phytoplankton size proxy (S_f) is derived by weighting two hyperspectral eigen vectors to the phytoplankton absorption spectrum, the vectors corresponding to populations of phytoplankton dominated by small cells or by large cells. *Devred et al.* (2011) (DSSP11) proposed a similar method limiting its absorption spectrum to a subset of wavelength specific to Ocean Color satellites. *Fujiwara et al.* (2011) (FUJI11) further simplified the method by using ratios of phytoplankton absorption bands from different pigments ($a_\varphi(443)/a_\varphi(667)$ and $a_\varphi(488)/a_\varphi(555)$. *Roy et al.* (2011) and *Roy et al.* (2013) (ROY11 and ROY13) took advantage of the pigment packaging by comparing the absorption peak at 676 with the chlorophyll *a* concentration to derive the median phytoplankton size (d) and the slope of the PSD (m_φ). Note that in my implementation of ROY11 and ROY13 like the other methods, I restrained the algorithm to information contained in the IOP and therefore used a chlorophyll *a* concentration derived from the IOP. This could impact the performance of the latter algorithm.

Methods presented up to now only leveraged the phytoplankton absorption spectrum, the anomalous dispersion method, also referred as the absorption band effects, proposed by *Houskeeper et al.* (2020) (HSKPR20), relies on the particulate beam attenuation to retrieve information on chlorophyll absorption and phytoplankton size. Here, I used five of the six eigen vectors (P1+, P1-, P2+, P3+, and P3-) derived by HSKPR20 from the beam attenuation residuals in combination with the PSD to derive size indexes of phytoplankton $D50_{\bar{G}}$ and m_φ. Note that these were computed by regressing them with our data (HSKPR20 did not have PSDs to compare to). The multivariate regression was computed with Matlab Regression Learner Toolbox with Cross-Validation enabled to protect against overfitting by partitioning the dataset into five folds.

The spectral shape of particulate beam attenuation (γ_{c_p}) and particulate backscattering ($\gamma_{b_{bp}}$) are known to covary with the mean size of suspended particles as developed in *Boss et al.* (2001a) (BOSS01) and used in *Fujiwara et al.* (2011). *Kostadinov et al.* (2009)

(KSM09) pushed the later by establishing an empirical relation between $\gamma_{b_{bp}}$) and m_{φ}. While these size indexes are for all particles they are expected to be close to the size indexes specific to phytoplankton as phytoplankton contributes to a major part of the particulate attenuation ($\sim$ 73 %) and backscattering signals (c.f. Chapter 3).

Another strategy explored to estimate a size index of the phytoplankton population was based on the ratio of backscattering to total scattering (b_{bp}/b_p): larger particles forward scatter relatively more than smaller particles (*Twardowski et al.*, 2001; *Boss et al.*, 2007a). This relation is expected to perform well with population of particles dominated by small particles. However, when the population of particles is dominated by larger particles, the composition of particles is expected to dominate b_{bp}/b_p.

2.3 Results and Discussion

2.3.1 Seasonal and Spatial Patterns in the Phytoplankton Size Distribution

In the western North Atlantic, phytoplankton size distribution varied significantly with seasons and subregions (Figure 2.1 & 2.15). Over the entire area of study the annual cycle could be summarized as follow: (1) low phytoplankton concentration at the surface during winter over the entire size spectrum (November); (2) a significantly higher concentration during spring (May) with larger cells ($>$ 7 µm) representing a high proportion of the total biovolume; (3) at the end of the summer concentrations of phytoplankton are still high compared to those in early and late winter, however the population tends to shift toward smaller cells with respect to spring (pico-plankton $<$ 2 µm in NAST and NATM subregions, and small nano-plankton 2 to 6 µm in ARCT area). In the NAST and GSSS subregions, the concentration of pico-plankton (dominated by Synechococcus) is high and stable throughout the year. In the ARCT subregion cells tend to be larger: large nano-phytoplankton contribute significantly to the total biovolume and the peak of pico-phytoplankton observed in the NAST and GSSS subregions is absent.

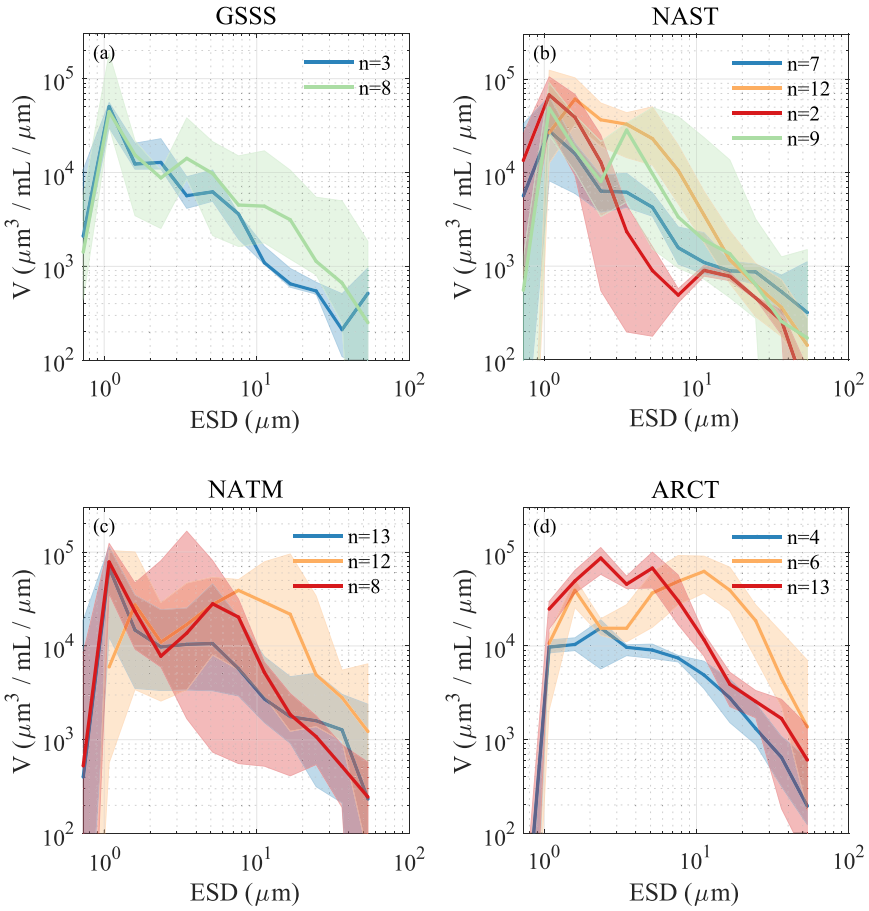

Figure 2.15. Phytoplankton size distribution (PSD) grouped by region and season. The transparent area surrounding each line correspond to the minimum and maximum PSD of each group. The subregions are from South to North: (a) the Gulf Stream and Sargasso Sea province (GSSS), (b) the North Atlantic Subtropical Gyral province (NAST), (c) the North Atlantic Temperate province (NATM), and (d) the Atlantic Arctic province (ARCT). PSD are colored by seasons: November (blue), March-April (green), May (orange), and September (red).

The predominant phytoplankton size groups observed throughout the western North Atlantic did not correspond to the traditional definition of phytoplankton size classes (pico < 2 µm, nano 2-20 µm, and micro > 20 µm *Sieburth et al.*, 1978). Historically size classes of phytoplankton were based on operational definition: the size of filter. For the size metric ESD, used here, the PSD grouping by seasons and subregions (Figures 2.15) and the sPLS analysis (Figure 2.16) suggested that phytoplankton population between 2-9 µm and

> 9 µm exhibited distinct relations to environmental parameters and therefore provide a more ecologically meaningful choice of cut-of size. The partition of size to make these groups is critically dependent on the metric used and care must be used when comparing size groups from different sources as the method used as a strong importance. For example, the difference in biovolume of micro-phytoplankton (> 20 µm) is greater than 20 % between the size metrics ESD and major axis (Supplemental material in *Bolaños et al.* (2020)).

The seasonal and geographical variability in the phytoplankton size distribution could be related to key environmental variables with a sparse partial least square analysis (sPLS, Figure 2.16). The mean Brunt-Väisälä frequency ($\bar{N}^2$), an indicator of the stability of the water column from the surface to 100 m, was positively correlated with the concentration of small nano-phytoplankton (2 to 6 µm). Larger phytoplankton (9 to 43 µm) where correlated with the concentrations of nutrients (Nitrate+Nitrite, Phosphate, and Silicate). These two parameters ($\bar{N}^2$ and nutrient concentration) have been widely used to predict the succession of phytoplankton communities in unstable environments (*Margalef*, 1978), here I confirm that they are relevant for such approach. Another feature revealed by the sPLS was the gradient in correlation between phytoplankton size, temperature and latitude, suggesting that larger phytoplankton are predominant at higher latitudes while pico-phytoplankton dominate in stations of the NAST and GSSS subregions. This observations is in agreement with part of the literature suggesting that in warmer waters smaller cells dominate and in cooler water the contribution of larger cells is enhanced (*Li et al.*, 2009; *Morán et al.*, 2010; *Hilligsøe et al.*, 2011). However, like *Marañón* (2015) I found exceptions to this pattern with large cells in the NAST subregion (in May) and small cells in the ARCT subregion (in September).

While the seasonal structure in size observed was well correlated with environmental parameters observed and modelled one aspect was not considered: the grazing communities. The annual cycle of phytoplankton as explained by the disturbance recovery hypothesis suggests that the seasonal structure in size could also be associated with

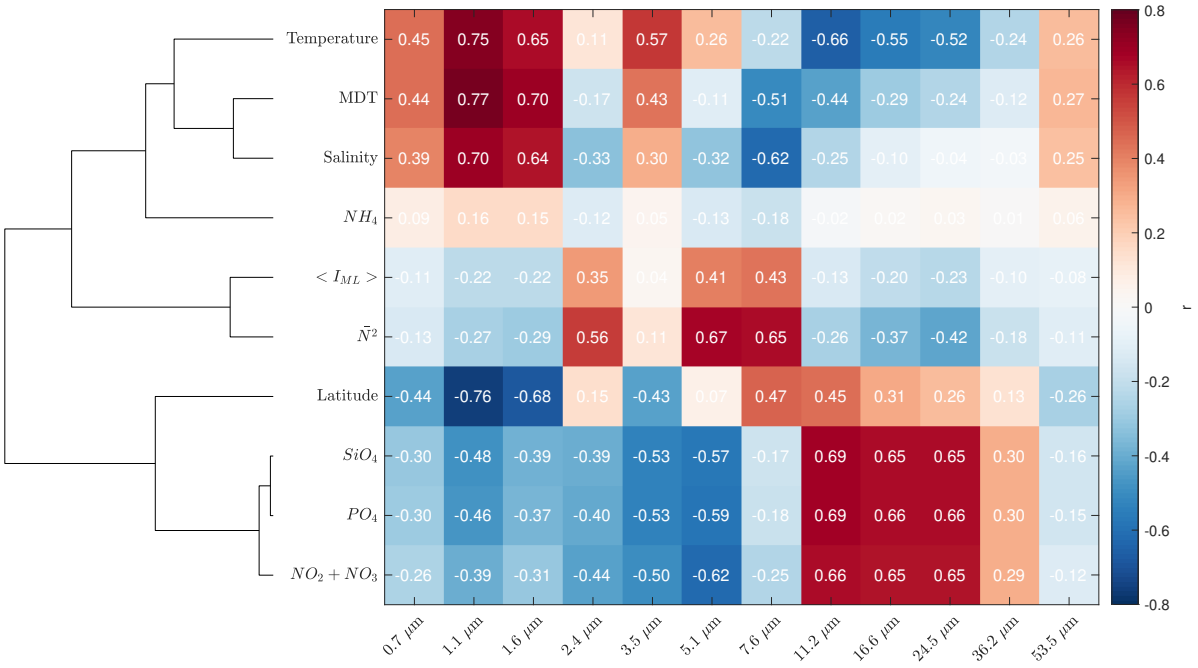

Figure 2.16. Phytoplankton sizes are correlated with environmental parameters as revealed by a sparse partial least squares (sPLS) analysis, a classical method for regression-based modeling. Correlation between size classes and environmental parameters are depicted as a clustered heat map. The normalization of the environmental data tried (none, relative, and zscore) had no impact on the output, and the normalization of the size classes (none or by the total biovolume) slightly changed the statistics but not the general trend. The sPLS was computed with the R package mixOmics (Rohart et al. 2017). The clustering was applied only on the rows with complete linkage algorithm and Euclidean distance metric.

different grazing communities (*Behrenfeld and Boss*, 2018). In the ARCT and NATM subregions, the deep winter mixing dilutes phytoplankton which is consistent with the low phytoplankton concentrations observed at every size class at the surface (Figure 2.15.c-d, blue lines). This dilution results in decreased light availability and encounter rates (predator-prey interactions). In spring, larger phytoplankton (9 to 43 µm) dominated the community while the mixed layer shallowed ((Figure 2.15.c-d, orange lines, Figure 2.1). This is in agreement with the decoupling between predator and prey hypothesis (*Kiørboe* (1993); *DeLong and Vasseur* (2012)). During the summer months, the phytoplankton assemblages were dominated by smaller cells (2 to 6 µm, Figure 2.15.c-d, red lines) which

could be explained by the water column stratifying, the surface layer being depleted in nutrients, and grazers catching up to their prey (*Chisholm*, 1992; *Kiørboe*, 1993). At the stations in the NAST and GSSS subregions, smaller cells were observed throughout the year (Figure 2.15.a-b). The environment was relatively stable throughout the year (stable MLD), conditions favorable of smaller cells, as they are more efficient at harvesting light and uptaking nutrients (*Chisholm*, 1992; *Kiørboe*, 1993; *Tremblay et al.*, 2009).

To my knowledge, these results present the first phytoplankton size distribution extended over a broad size range (0.6 to 65 µm) in the western North Atlantic. It confirms ecological theories that were mainly validated with particulate size distribution. In addition, it supplements a scarce dataset of phytoplankton size distributions (*Huete-Ortega et al.*, 2011; *Laney and Sosik*, 2014; *Marañón*, 2015), compared to the many more particulate size distributions measured.

2.3.2 Evaluation of Optical Size Indexes

The phytoplankton size distribution built here presents a unique opportunity to test phytoplankton size index algorithms developed to date, which are applied to Ocean Color satellites (*Mouw et al.*, 2017). I first compare the spatial and temporal patterns in size structure observed with the PSD (previous section) to the patterns retrieved with a subset of optical size indexes. I then evaluate a multitude of optical size index algorithm based on IOPs.

The size structure patterns across the western North Atlantic described by optical size indexes were similar to the patterns observed with the PSD for algorithms based on biomass, relative pigment concentrations, absorption band effect, and spectral slope (Figure 2.17). In most cases, the temporal size structure within a region and in between regions was clearly discriminated. However, the complexity of the PSD cannot be synthesized by one parameter, for example in the NAST region the PSD in May has significantly more nanophytoplankton ranging from 3 to 7 µm than in September. This was

not captured by the spectral slope method BOSS01 but was observed by both relative pigment concentrations and absorption band effect methods (CB06 and HSKPR20).

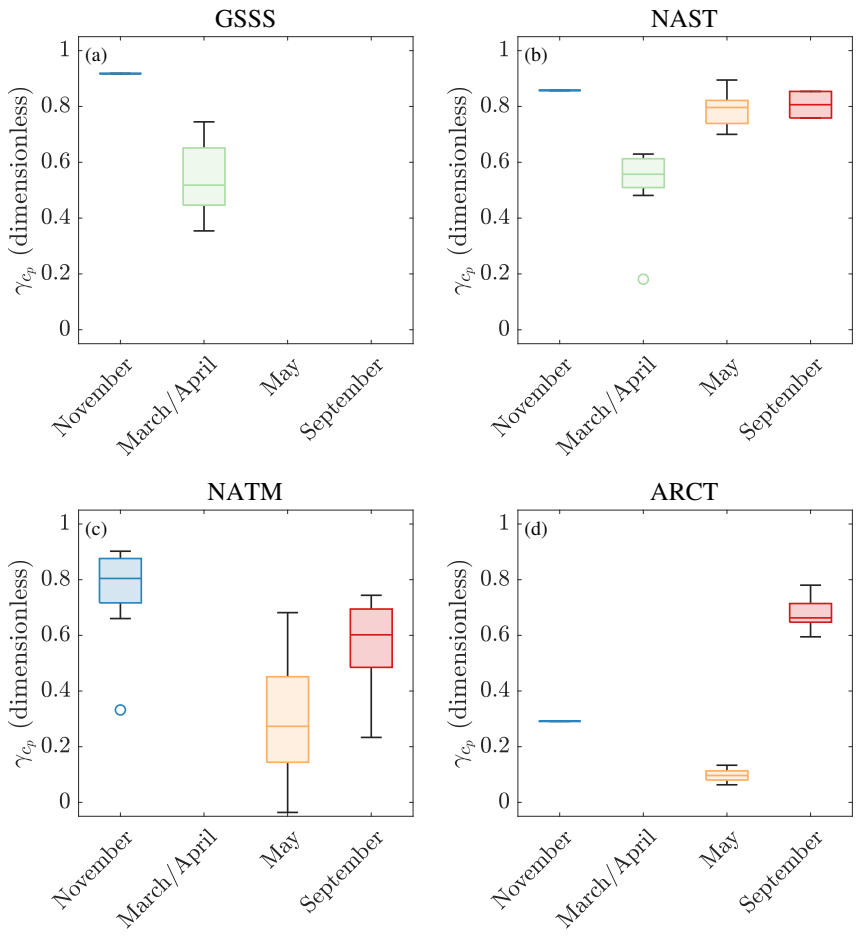

Figure 2.17. Seasonal and spatial size structure observed with the optically derived index γ_{c_p} based on the spectral slope of beam attenuation. The subregions are from South to North: (a) the Gulf Stream and Sargasso Sea province (GSSS), (b) the North Atlantic Subtropical Gyral province (NAST), (c) the North Atlantic Temperate province (NATM), and (d) the Atlantic Arctic province (ARCT). On each box plot, the central mark indicates the median, the top and bottom edges indicate the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentile. The whiskers extend to the most extreme data points excluding outliers (denoted by the 'o' symbol).

The statistical relation between each size index method and the metrics derived from the PSD (m_φ, $D50_N$, $D50_{\bar{G}}$, and $D50_V$) varied greatly and are presented in Figure 2.18 and Table 2.4. The highest correlations were found with $D50_{\bar{G}}$ consistent with the expectation that optical properties in the size range of phytoplankton primarily depends on

the cross-sectional area. The methods based on relative pigment concentration and packaging correlated to some extent, our results suggest that among these methods the ones utilizing hyperspectral data are good predictors (CB06). *Fujiwara et al.* (2011) suggest that the limited correlation could be explained by the influence of the pigments of specific taxons on the proxies developed (e.g. small diatoms have the same pigments as larger ones). The most promising phytoplankton size proxies were based on the spectral slope of particulate attenuation (γ_{c_p}) and absorption band effect with a correlation coefficient higher than 0.82 and median absolute relative offset (MARD) as low as 7 %. Note that the size index derived with the absorption band effect method (HSKPR20) was partially built with the present dataset potentially enhancing its performance. An encouraging algorithm with an MARD of 8 % and correlation of 0.46 was the spectral slope of backscattering ($\gamma_{b_{bp}}$). However, $\gamma_{b_{bp}}$ was derived from only three wavelengths (470, 532, and 650 nm), potentially reducing its performance. Nonetheless, these results suggest that the spectral shape of backscattering could be used for further development of size algorithm. The recent development of hyperspectral backscattering sensors (Sequoia Hyper-bb) in combination with the future deployment of hyper spectral satellite sensor (NASA PACE) will enhance our capacity to infer phytoplankton size at a global scale.

To my knowledge, these results are the first independent validation of optical size indexes with direct measurement of phytoplankton size. In fact, most of these algorithms have only been tested against diagnostic pigment analysis (DPA, *Uitz et al.*, 2006) which is only a proxy for phytoplankton size. The DPA was only recently compared to an independent set of size measurement and significant offsets were reported (*Chase et al.*, 2020).

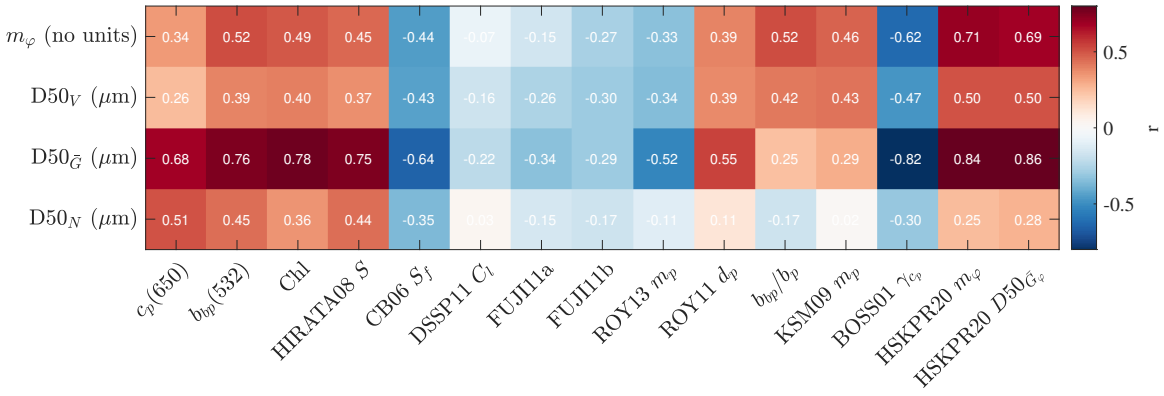

Figure 2.18. Heatmap of correlation coefficients between phytoplankton size distribution (PSD) metrics and phytoplankton size proxies derived from the inherent optical properties (IOPs). The metrics of the PSD are its slope (m_φ, first column) and the median diameter based on volume, average cross-sectional area, and abundance ($D50_N$, $D50_{\bar{G}}$, and $D50_V$, second to last columns). The optical algorithms are particulate attenuation at 650 nm ($c_p(650)$), particulate backscattering at 532 nm ($b_{bp}(532)$), chlorophyll a concentrations derived from line height (Chl, µgL^{-1}), the size index (S, dimensionless) from Hirata et al. 2008 (HIRATA08), the size index S_f (dimensionless) from Ciotti and Bricaud et al. 2006 (CB06), the contribution of large phytoplankton to the total chlorophyll concentration (C_l, %) from Dreved et al. 2011 (DSSP11), the ratios of phytoplankton absorption ($a_\varphi(488)/a_\varphi(555)$) from Fujiwara et al. 2011 (FUJI11b), the average phytoplankton population diameter (d, µm) from Roy et al. 2011, (ROY11) the PSD slope (m_φ) derived from absorption with Roy et al. 2013 (ROY13), the ratio of particulate backscattering to scattering (b_{bp}/b_p), the PSD slope (m_φ) derived from particulate backscattering with Kostadinov et al. 2009 (KSM09), the size index γ_{c_p} based on the spectral slope of beam attenuation from Boss et al. 2001 (BOSS01), the PSD slope (m_φ) and the median average cross-sectional area ($D50_{\bar{G}}$) derived from Houskeeper et al. 2020 (HSKPR20). The dashed black line correspond to a robust least square regression. Statistics of relations are presented Table 2.4.

Table 2.4. Statistics of relations between phytoplankton size distribution (PSD) metrics and phytoplankton size proxies derived from the inherent optical properties (IOPs). The metrics of the PSD were its slope (m_φ) and the median diameter based on volume, average cross-sectional area, and abundance ($D50_N$, $D50_{\bar{G}}$, and $D50_V$). The Pearson's correlation coefficient (r) was computed directly between each metric. The median absolute deviation and median absolute relative deviation were computed between the PSD metric and the PSD metric predicted from the IOP algorithm using a robust linear least square regression. Bold numbers correspond to algorithms performing better than the chlorophyll a concentration for that metric.

PSD metric	IOP algorithm	r	MAD	MARD
m_φ	Chl	0.49	0.35	0.08
m_φ	$b_{bp}(532)$	0.51	0.42	0.10
m_φ	$c_p(650)$	0.33	0.44	0.11
m_φ	HIRATA08	0.45	0.39	0.09
m_φ	CB06	-0.44	0.44	0.12
m_φ	DSSP11	-0.06	0.48	0.12
m_φ	FUJI11a	-0.15	0.48	0.11
m_φ	FUJI11b	-0.26	0.50	0.12
m_φ	ROY11	0.38	0.39	0.09
m_φ	ROY13	-0.33	0.45	0.11
m_φ	b_{bp}/b_p	**0.52**	**0.31**	**0.08**
m_φ	KSM09	0.46	0.30	**-0.08**
m_φ	BOSS01	**-0.62**	**0.24**	**0.06**
m_φ	HSKPR20 m_φ	**0.70**	**0.28**	**0.06**
m_φ	HSKPR20 $D50_{\bar{G}}$	**0.68**	**0.29**	**0.07**
$D50_V$	Chl	0.40	1.11 µm	0.29
$D50_V$	$b_{bp}(532)$	0.39	1.50 µm	0.30
$D50_V$	$c_p(650)$	0.26	1.73 µm	0.36
$D50_V$	HIRATA08	0.36	1.42 µm	**0.28**
$D50_V$	CB06	**-0.43**	2.15 µm	0.34
$D50_V$	DSSP11	-0.16	3.16 µm	0.46
$D50_V$	FUJI11a	-0.25	2.93 µm	0.43
$D50_V$	FUJI11b	-0.29	2.62 µm	0.38
$D50_V$	ROY11	0.38	2.50 µm	0.33
$D50_V$	ROY13	-0.34	2.40 µm	0.37
$D50_V$	b_{bp}/b_p	**0.42**	2.29 µm	0.36
$D50_V$	KSM09	**0.43**	2.08 µm	**0.28**
$D50_V$	BOSS01	**-0.47**	**0.84** µm	**0.17**
$D50_V$	HSKPR20 m_φ	**0.50**	1.31 µm	**0.28**
$D50_V$	HSKPR20 $D50_{\bar{G}}$	**0.50**	**1.03** µm	**0.22**
$D50_{\bar{G}}$	Chl	0.78	0.44 µm	0.23
$D50_{\bar{G}}$	$b_{bp}(532)$	0.75	0.96 µm	0.40
$D50_{\bar{G}}$	$c_p(650)$	0.67	0.45 µm	0.29
$D50_{\bar{G}}$	HIRATA08	0.75	0.63 µm	0.29
$D50_{\bar{G}}$	CB06	-0.64	1.05 µm	0.44
$D50_{\bar{G}}$	DSSP11	-0.21	0.98 µm	0.50
$D50_{\bar{G}}$	FUJI11a	-0.33	1.08 µm	0.48
$D50_{\bar{G}}$	FUJI11b	-0.28	1.10 µm	0.51
$D50_{\bar{G}}$	ROY11	0.54	1.19 µm	0.55
$D50_{\bar{G}}$	ROY13	-0.52	1.17 µm	0.54
$D50_{\bar{G}}$	b_{bp}/b_p	0.24	1.08 µm	0.50
$D50_{\bar{G}}$	KSM09	0.29	1.04 µm	0.44
$D50_{\bar{G}}$	BOSS01	**-0.81**	0.65 µm	0.25
$D50_{\bar{G}}$	HSKPR20 m_φ	**0.83**	0.97 µm	0.41
$D50_{\bar{G}}$	HSKPR20 $D50_{\bar{G}}$	**0.86**	0.90 µm	0.40
$D50_N$	Chl	0.35	0.18 µm	0.16
$D50_N$	$b_{bp}(532)$	**0.44**	**0.14** µm	**0.13**
$D50_N$	$c_p(650)$	**0.50**	**0.15** µm	**0.14**
$D50_N$	HIRATA08	**0.43**	**0.16** µm	**0.14**
$D50_N$	CB06	**-0.35**	0.21 µm	0.17
$D50_N$	DSSP11	0.03	0.22 µm	0.20
$D50_N$	FUJI11a	-0.14	0.22 µm	0.19
$D50_N$	FUJI11b	-0.17	0.22 µm	0.19
$D50_N$	ROY11	0.11	0.22 µm	0.20
$D50_N$	ROY13	-0.11	0.22 µm	0.20
$D50_N$	b_{bp}/b_p	-0.17	0.20 µm	0.19
$D50_N$	KSM09	0.02	0.23 µm	0.20
$D50_N$	BOSS01	-0.30	0.19 µm	0.17
$D50_N$	HSKPR20 m_φ	0.25	0.21 µm	0.19
$D50_N$	HSKPR20 $D50_{\bar{G}}$	0.27	0.21 µm	0.19

2.3.3 Phytoplankton Size Distribution Closure

Phytoplankton size distribution of the NAAMES campaigns, presented Figure 2.10.d-f, are evaluated against independent observations relating to size: chlorophyll *a* concentration, inherent optical properties (IOP), and the particulate size distribution. In addition, the PSD from both cytometers are compared to each other in their overlapping size range.

The number of particles counted by the ICS and the IFCB in their overlapping interval ($\sim$ 7.1 to $\sim$ 9.3 µm) was similar with no significant offset for an instrument with the exception of the third NAAMES campaign for which the number of particles counted by the ICS was greater than the IFCB (Figure 2.19.h). The instruments were off by less than a factor of two, which is low considering the seven orders of magnitude spanned by the abundance PSD. The offset from the third campaign is likely due to the IFCB underestimating the number of small particles. This discrepancy can be explained by the low chlorophyll *a* concentration per cell pertaining to that campaign (Figure 2.19.d) causing cells to fluoresce less and resulting in some cells going undetected by the IFCB.

Total phytoplankton volume ($\mu m^3 mL^{-1}$) derived from each PSD was converted to chlorophyll *a* concentrations and compared to HPLC as in *Chase et al.* (2020). The only difference with *Chase et al.* (2020) was the updated calibration of the PSD here. The chlorophyll *a* from both PSD and HPLC covaried as expected ($r^2 = 0.78$) with reasonable differences from a linear relationship in both directions (Figure 2.19.d). The variability in the relation is expected and can be explained by abiotic factors (e.g. light level, nutrient concentrations, MLD). At the end of the summer (August-September), the cells are expected to be high light acclimated and in low nutrient concentrations due to the shallow MLD, conditions in which a low ratio of chlorophyll to cell volume is expected and observed in our dataset.

Several optical proxies were used to establish closure with the PSD. Particulate attenuation and backscattering correlated well with phytoplankton total cross-sectional area computed from the PSD ($r^2 = 0.81$ and 0.60 respectively, Figure 2.19.b-c) as expected

(e.g. *Behrenfeld and Boss*, 2006; *Dall'Olmo et al.*, 2009). In addition, the efficiency of phytoplankton to absorb or attenuate light was computed and compared to established values throughout the literature. The absorption efficiency factors ($Q_{a_\varphi}(676) = a_\varphi(676)/\Sigma\bar{G}_\varphi = 0.1 \pm 0.1$, Figure 2.19.a,e) from this study were slightly lower but within the uncertainties of the values from a laboratory experiment previously reported in the literature ($\sim 0.19 \pm 0.12$ for Morel and Bricaud 1986). The attenuation efficiency factors are typically in the range 1 to 3 according to Mie theory (*van de Hulst*, 1957). In the present study as I only considered the cross-sectional area of phytoplankton (not accounting for non-algal particles), I would expect lower values for the ratio $c_p/\Sigma\bar{G}_\varphi$ which is in agreement with our observations (Figure 2.19.c, more details in Chapter 3). This exercise revealed that the phytoplankton PSDs were consistent with the measured IOPs of the particles.

The power-law-fit exponents (m_φ) of the PSD expressed in abundance ($N(D)$) were in the same range as the ones derived from the particulate size distribution from the LISST (m_p, Figure 2.19.f) and consistent with the literature (3.7 to 4.7 on average according to 18 studies, Table 5.8 of *Jonasz and Fournier* (2007)). In addition, the size index γ_{c_p} was found to correlate well with the slope of the PSD (Figure 2.19.g). These results suggest that the non-algal particles, which dominate POC, co-vary in size with phytoplankton.

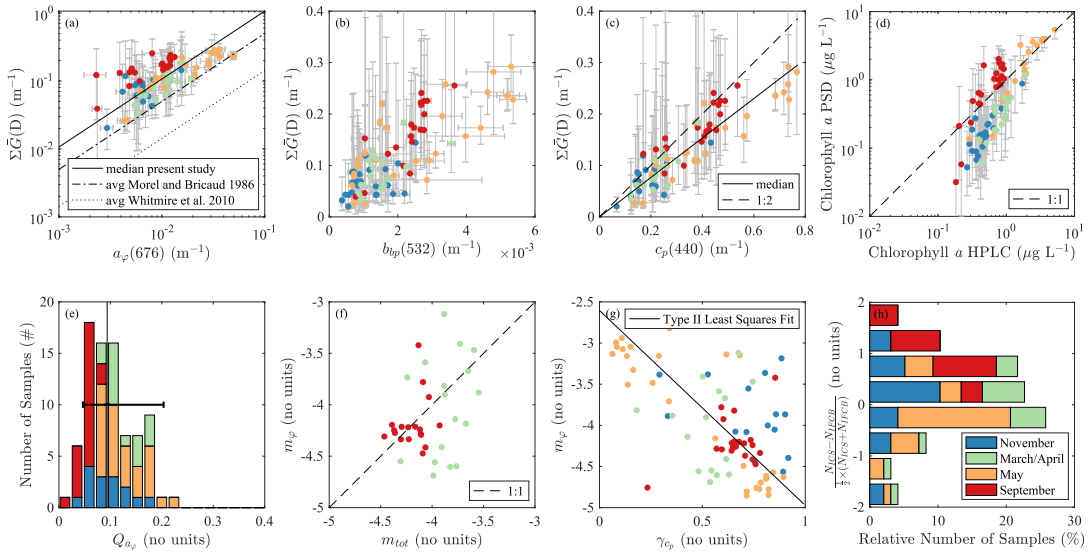

Figure 2.19. Assessment of the phytoplankton size distribution (PSD) from the NAAMES dataset. The total average cross-sectional area ($\Sigma\bar{G}$) is correlated to the inherent optical properties measured at the same time: phytoplankton absorption at 676 nm (a), particulate backscattering at 532 nm (b), and particulate attenuation at 440 nm (c). The phytoplankton absorption efficiency (Q_{a_φ}) by sample is presented in (e) with a median of 0.09 $\pm$ 0.03 ($\pm$MAD). (d) Relation between total phytoplankton volume from the PSD converted to chlorophyll a concentration (using published relations) and chlorophyll a concentration. (f) Relation between the PSD slope (m_φ) and the slope of the particulate size distribution (m_p) measured by the LISST. (g) Relation between the PSD slope (m_φ) and the size index parameter γ_{c_p}. (h) Relative difference between the ICS and IFCB in number of phytoplankton cells counted in their overlapping interval. The error bars for $\Sigma\bar{G}$, ΣV and Chlorophyll a derived from the PSD correspond to the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile of the Monte-Carlo uncertainty analysis of the PSD. The error bars of the IOPs correspond to the standard deviation of matching observations. The error bar of Q_{a_φ} correspond to the shift of the histogram using the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile of the Monte-Carlo uncertainty analysis in place of the value derived from the PSD.

2.4 Conclusion

In this work I derived phytoplankton size distributions for samples collected during four seasons in the western North Atlantic spanning from the subtropical to the subpolar gyre. To derive such size distribution with reasonable uncertainties, cytometers used must be well calibrated. Hence the importance of constantly checking the stability of the instrument's signals with reference beads ran before and after samples and a proper size

calibration by (1) using cultures of known size and optical properties or (2) in-situ cells sized by microscopy. The dataset built in the western North Atlantic is unique and contribute significantly to the scarce observations of phytoplankton size distributions which contrast with the well characterized particulate size distribution.

The trends in the size distribution observed align well with our current understanding of the phytoplankton annual cycle; PSDs are dominated by picoplankton in the North Atlantic subtropical gyre province throughout the year while having an increased contribution by small nano-phytoplankton (< 9 µm) in spring. In northern provinces (ARCT and NATM), a shift in size is observed through the year: from larger phytoplankton dominating the community in spring, transitioning to small nanophytoplankton throughout the summer, to a flatter size distribution associated with low concentrations early in the winter.

In addition, I tested algorithms based on five different strategies to obtain information on phytoplankton size. I found that most correlated less well with size than did chlorophyll. The absorption band effect and spectral slope of attenuation and scattering were the best predictors of phytoplankton size. This suggest that algorithm development effort associated with ocean color should focus on obtaining the backscattering spectra with better accuracy than currently done (*Bisson et al.*, 2019).

CHAPTER 3

PHYTOPLANKTON AND NON-ALGAL PARTICLES CONTRIBUTION TO THE INHERENT OPTICAL PROPERTIES OF THE SURFACE OCEAN, INSIGHTS FROM CYTOMETRY AND THE LISST PARTICLE SIZER

3.1 Introduction

Phytoplankton biomass is routinely estimated from optical measurements of the ocean. Autonomous vehicles (e.g. glider, floats, sail-drones) and ocean color satellite carrying optical sensors are the only methods of observations which provide a synoptic view of the ocean (*Bax et al.*, 2019; *Claustre et al.*, 2020). Phytoplankton properties are often derived empirically from the bulk optical properties of the ocean (*Behrenfeld et al.*, 2005; *Behrenfeld and Boss*, 2006; *Cetinić et al.*, 2012; *Boss et al.*, 2015; *Graff et al.*, 2015; *Rasse et al.*, 2017). Bulk optical properties are divided into two classes: inherent optical properties (IOP) when only depending on the medium (e.g. attenuation, backscattering) and apparent optical properties (AOP) when depending on both the medium and, to a smaller degree, the ambient light field (e.g. reflectance). The connection between the IOPs and AOPs is provided by the radiative transfer theory. The relationships between biogeochemical quantities and IOPs depends on our ability to differentiate the constituents (e.g. phytoplankton, non-algal particles, dissolved organic matter) that contributes to the bulk optical properties. Hence, to constrain phytoplankton properties with these observations it is essential that phytoplankton specific optical properties are measured and distinguished from other water constituents.

An analysis of light scattering budget in typical open ocean water suggested that only a small fraction of the backscattering (< 12 %) was explained by phytoplankton (*Brown and Gordon*, 1974; *Stramski and Kiefer*, 1991; *Morel and Ahn*, 1991; *Zhang et al.*, 2011). These budgets were constructed based on measurements of particles present in the water and an application of Mie theory (which assumed phytoplankton to be homogeneous spheres).

Hence, researchers speculated that sub-micron particles (0.4-1 µm) must be a major source of backscattering in the ocean (*Stramski and Kiefer*, 1991; *Zhang et al.*, 2011). More recently comparison between observations and the modeling of backscattering in the open ocean suggested that the contribution of phytoplankton or particles to the backscattering budget was significantly underestimated (*Volten et al.*, 1998; *Vaillancourt*, 2004; *Dall'Olmo et al.*, 2009; *Whitmire et al.*, 2010; *Harmel et al.*, 2016). This mismatch is referred to as the 'missing backscattering enigma' (*Stramski et al.*, 2004). A growing body of literature suggests that the missing backscattering stems from the structural complexity of phytoplankton and suggested using a coated sphere model for phytoplankton (*Organelli et al.*, 2018), as it increases backscattering by more than one order of magnitude (*Meyer*, 1979; *Kitchen and Zaneveld*, 1992). However, the quantification of the phytoplankton contribution to scattering remains a challenge due to the scarcity of concurrent measurements of optical properties and independent phytoplankton size distributions (e.g. microscopy or cytometry).

Here, I take advantage of the concurrent observations of the IOPs and phytoplankton size distribution (PSD) in the western North Atlantic collected as part of the North Atlantic Aerosol and Marine Ecosystem Study (NAAMES, *Behrenfeld et al.*, 2019). Methodologies used to construct the PSD from a combination of cytometers are described in chapter 2. The particulate efficiency factors, defined as the ratios of the radiative energy absorbed, attenuated, or backscattered by all the particles to the total geometrical cross-sectional area of these particles, was used in combination to constrain quantitatively the contribution of phytoplankton to particulate backscattering and hyperspectral particulate absorption and attenuation.

3.2 Methods

3.2.1 Field Campaigns and Data Processing

NAAMES campaigns targeted different phases of the annual cycle of phytoplankton in the western North Atlantic (40-58°N, 50-35°W). Sampling took place during November 2015 (autumn, NAAMES I), March 2018 (winter, NAAMES IV), May 2016 (spring, NAAMES II), and September 2017 (summer, NAAMES III). These campaigns covered a wide range of physical conditions and biogeochemical provinces (c.f. Chapter 2, [chla] $\in$ [0.05 5.14] µg/L and MLD $\in$ [5 255] m).

IOPs were measured continuously while underway with a flow through system pumping water from ∼5 m below the surface with a diaphragm pump to prevent damaging the particles (*Cetinić et al.*, 2016). A WetLabs ACS measured hyperspectral absorption and attenuation of light (25 cm path lengths, 400-750 nm), a WetLabs ECO-BB3 measured backscattering at three wavelengths (470, 532, and 650 nm), and a Sequoia LISST Type B (670 nm) measured forward scattering. The LISST was only used during the last two campaigns. Instruments were setup and calibrated following the recommendations of *Boss et al.* (2019) and the calibration independent technique of *Slade et al.* (2010) to obtain particulate absorption (a_p), attenuation (c_p), and backscattering (b_{bp}). In NAAMES III and IV the particulate size distribution of the total particle pool >2 µm was estimated from the forward scattering by the LISST (similar to *Boss et al.*, 2018b).

The cytometers, a BD Influx Cell Sorter flow cytometer (ICS) and a McLane Imaging Flow CytoBot (IFCB)), were operated on discrete samples collected from the underway system or from surface niskin bottles; or continuously (IFCB only) from the underway system. For both instruments, only particles fluorescing at a red wavelength (sensitive to chlorophyll *a*) beyond a set threshold were used and assumed to be phytoplankton cells for the present study ($\sim$ 8 % uncertainty in PSD, c.f. Chapter 2). Size calibration, retrieval of the PSD, and merging of the PSD from both instruments was detailed in the Chapter 2. The average volume sampled per matchup with the optics measurements was 0.8 mL for

the ICS (lower end of the size spectrum <10 µm) and 29 mL for the IFCB (>7 µm). IOP observations were averaged on a window of an hour and a half around the time of collection of the PSD sample unless a strong variation in chlorophyll *a* concentration – indicating a possible change in water mass – was noted, in which case a smaller time window was used. The procedure resulted in 82 samples with the ACS and 91 samples with ECO-BB3 (the difference is due to times when data from one instrument that could not be recovered due to bubbles).

3.2.2 Efficiency Factors

The approaches to compute the contribution of phytoplankton to particulate attenuation, absorption, and backscattering rely on the computation of the efficiency factors, as these are well constrained by optical theory. The particulate efficiency factors for attenuation (Q_{c_p}), backscattering ($Q_{b_{bp}}$), and attenuation (Q_{a_p}) are defined by the ratios of particulate optical properties to the total cross sectional area of particles:

$$Q_{c_p} = \frac{c_p}{G_p}, \quad Q_{b_{bp}} = \frac{b_{bp}}{G_p}, \quad Q_{a_p} = \frac{a_p}{G_p} \tag{3.1}$$

where G_p represents the total particulate cross-sectional area projected to a plane perpendicular to the light beam, c_p the particulate attenuation, b_{bp} the particulate backscattering, and a_p the particulate absorption. Similarly, the phytoplankton efficiency factors (Q_{c_φ}, $Q_{b_{b\varphi}}$, Q_{a_φ}) are computed replacing particulate parameters by phytoplankton parameters (c_φ, $b_{b\varphi}$, a_φ, G_φ).

3.2.3 Computation of the Phytoplankton Absorption Efficiency Factor

To compute the absorption efficiency factor of phytoplankton (Q_{a_φ}), both the phytoplankton absorption (a_φ) and the phytoplankton cross-sectional area are required. Three methods were used to estimated a_φ: (1) the absorption line height method from *Boss et al.* (2007b) was used for $a_\varphi(\lambda = 676)$; (2) the spectral decomposition method of *Chase et al.* (2013) using vector specific to each phytoplankton pigment was used to

obtained $a_\varphi(\lambda)$ at all wavelengths; (3) the method of *Zheng and Stramski* (2013) separating the absorption of phytoplankton from non-algal particles (NAP) based on stacked-constraints was used to obtained $a_\varphi(\lambda)$ at all wavelengths.

The total phytoplankton cross-sectional area (G_φ) was computed by integrating the phytoplankton size distribution obtained from the cytometers. Here I assumed that the particles were randomly oriented when flowing in front of the optical sensors, therefore I used the average cross-sectional area for phytoplankton sized with the cytometers. The average cross-sectional area was estimated from the surface area of the particles (based on *Moberg and Sosik*, 2012) using Cauchy's theorem that state: the mean projected area of a randomly oriented convex particle is one quarter of its surface area (*Cauchy*, 1832; *Vickers and Brown*, 2001). This assumption did not have a significant impact on the total cross-sectional area retrieved from the PSD as only a few larger taxons were found to have a difference between their observed cross-sectional area and the cross-sectional area computed from the surface area (c.f. Chapter 2).

3.2.4 Contribution of Phytoplankton to Particulate Attenuation

Because the particulate optical properties are not only due to phytoplankton, the contribution of phytoplankton to particulate attenuation (c_φ/c_p) was estimated from the measured particulate attenuation (c_p) and the measured total cross-sectional area of phytoplankton (G_φ). The phytoplankton attenuation (c_φ) was retrieved from G_φ and the efficiency factors of phytoplankton (Q_{c_φ}) which was computed with optical theory (*Mie*, 1908; *van de Hulst*, 1957).

A consequence of Babinet's principle is that the light attenuated from a beam due to a particle that is significantly larger than the wavelength is equal to twice the particle's cross section ($Q_c = 2$), known as the extinction paradox (*van de Hulst*, 1957). In fact, the efficiency factor of particulate attenuation computed for a population of randomly oriented particles varying in size, shape and index of refraction representative of phytoplankton

asymptotes to two as mean particle size increases while typically ranging between 0.5 and 3 for smaller cells (> 1 µm) with tight size distribution (see *Morel and Bricaud* (1986) for homogeneous spheres, *Clavano et al.* (2007) for spheroids, *van de Hulst* (1957), Figure 72 and 73 for homogeneous non-spherical particles). Here, I used Mie theory that is assuming all particles are homogeneous spheres, as described in *Bohren and Huffman* (1983), to compute $Q_{c_\varphi}(\lambda)$ specific to each phytoplankton size distribution. While modelled absorption, total scattering, and beam attenuation of marine particles and phytoplankton were in good agreement with observations, this is not the case for backscattering (*Stramski et al.*, 2001). Mie theory takes as input the relative index of refraction of the particles with respect to the medium (m_r) and the size parameter x which depends on the particles size (D), the index of refraction of the medium (m_m), and the wavelength (λ).

$$m_r = \frac{m_p + im'_p}{m_m} \tag{3.2}$$

$$x = \frac{\pi D m_m}{\lambda} \tag{3.3}$$

Following *Organelli et al.* (2018), I used a real index of refraction $m_p = 1.06$ and an imaginary index of refraction $m'_p \in [0.010.0003]$. I ran Mie theory for each size bin of the PSD to compute the efficiency factor of each particles size (Q_{c_i}). To approximate the efficiency factor of the phytoplankton population (Q_{c_φ}), I computed an area-based weighted average of Q_{c_i} as follow:

$$Q_{c_\varphi}(\lambda) = \sum_{D=D_{min}}^{D_{max}} Q_{c_i}(D,\lambda) \times \frac{G_\varphi(D)}{\sum_{D=D_{min}}^{D_{max}} G_\varphi(D)} \tag{3.4}$$

With D_{min} and D_{max} the minimum and maximum sizes of the size distribution respectively (0.6 µ µm and 65 µm, full size bin description in Chapter 2). The ratio of phytoplankton to particulate attenuation could simply be computed by estimating the phytoplankton attenuation with the attenuation efficiency factor equation (Eq. 3.1).

To establish a closure of the optical model used and improve our understanding on which particle size contributes most to the particulate attenuation, the measured total

cross-sectional area of particles ($G_p = G_\varphi + G_{NAP}$) was used in combination of the same optical model to compute the particulate attenuation.

Another approach to estimate the contribution of phytoplankton to c_p was to compare the total particulate and phytoplankton cross-sectional area (G_p and G_φ). The particulate size distribution was obtained from the LISST after inverting the forward scattering measurements with the manufacturer's Matlab software as in *Boss et al.* (2018b). The calibration of the particulate size distribution obtained depends on a calibration constant (VCC) which is provided by the manufacturer and typically computed with measurements of a solution of known particulate concentration. Here I further checked the calibration by reconstructing the particulate attenuation ($c_{p_{Mie}}$) from the particulate size distribution and compared it with the beam attenuation measured by the LISST ($c_{p_{obs}}$).

I first extended the particulate size distribution to the same size range as the IOP observations (down to 0.2 µm, the size of the filter we use to separate dissolved from particulate measurements). A power law with the exponent of the observed particulate size distribution was used to extend the abundance of particles beyond 2.05 µm to 0.6 µm. The abundance extrapolated at 0.6 µm was then set for all smaller sizes (from 0.6 to 0.2 µm); saturating the particle abundance below 0.6 µm. The three first bins (<2.05 µm) of the LISST's size distribution were not used per recommendations of the literature (*Traykovski et al.*, 1999) as the LISST inversion 'folds' in to smaller bins, the effect of particles smaller than the inversion computes is of greater importance on the forward scattering.

I then computed the attenuation efficiency factor (Q_{c_i}) for each size bins of the LISST, using the same Mie procedure described above with a set of indices of refraction ($m_p \in [1.05, 1.1]$, $m'_p \in [0.01, 0.001]$) representative of phytoplankton and NAP (*Aas*, 1996). The particulate attenuation ($c_{p_{Mie}}$) can then be retrieved as follow:

$$c_{p_{Mie}}(670) = \sum_{D=D_{min}}^{D_{max}} Q_{c_i}(D, 670) \times G_p(D) \tag{3.5}$$

With $D_{min} = 0.2$ µm and $D_{max} = 230$ µm. The particulate attenuation retrieved from the particulate size distribution ($c_{p_{Mie}}$) was close to the particulate attenuation directly

measured by the LISST ($c_{p_{obs}}$). In fact, to force them to match the VCC needed to be adjusted by a factor of 4 % (0.96), ranging between 0.64 and 1.25 for the set of indexes of refraction tested. These suggest that the VCC provided by the manufacturer and specific to our instrument was consistent with the measurements and with the conclusions of (*Boss et al.*, 2018b).

3.2.5 Contribution of Phytoplankton to Particulate Backscattering

Both the contribution of phytoplankton to particulate backscattering and the phytoplankton's backscattering efficiency factors are required to estimate phytoplankton backscattering ($b_{b\varphi}$).

The pathway adopted here takes advantage of the recent findings of *Organelli et al.* (2018), suggesting that particulate backscattering can be predicted from a coated sphere optical model. I used the coated sphere model implementation from (*Zhang et al.*, 1998) with a realistic set of parameters from Figure 3 of *Organelli et al.* (2018) to retrieve the individual phytoplankton and particulate backscattering efficiency factors ($Q_{b_{bpi}}$ and $Q_{b_{b\varphi i}}$). Bulk efficiency factors ($Q_{b_{bp}}$ and $Q_{b_{b\varphi}}$) were computed as for attenuation (Eq. 3.4). The bulk particulate backscattering efficiency was used to assess the model with observations while the bulk phytoplankton efficiency factor was used to estimate the contribution of phytoplankton to backscattering.

3.3 Results and Discussion

3.3.1 Phytoplankton Absorption Efficiency Factor

The hyperspectral absorption efficiency factor of phytoplankton (Q_{a_φ}) was derived for the western North Atlantic (Figure 3.1). The median value observed (0.15 ranging from 0.03 to 0.35 at 676 nm and 0.35 ranging from 0.1 to 0.74 at 442 nm) was slightly lower than laboratory experiments previously reported in the literature but well within the range of observed values. *Morel and Bricaud* (1986) observed a median $Q_{a_\varphi}(676)$ of 0.2 for 22

cultures ranging from 0.032 for high light acclimated Cyanobacteria to 0.459 for low light acclimated Isochrysis galbana. *Whitmire et al.* (2010) median $Q_{a_\varphi}(442)$ was 0.7 with values ranging from 0.24 to > 1. Variations in efficiency factors are due to light and nutrient conditions as observed in laboratory experiments from *Morel and Bricaud* (1986). In fact, when chlorophyll concentration per cell was low (in September) the absorption per cell was lower: they were more efficient at absorbing light (less self-shading) but less light was absorbed per cross-sectional area, hence lowering their absorption efficiency factor (Figure 3.2).

Methods to retrieve a_φ were consistent across all samples; however, a bias was noted. The line height method tended to be 47 % lower than the spectral decomposition method at 676 nm. On one hand, the line height method from *Boss et al.* (2007b) was more likely to underestimate a_φ as the 676 chlorophyll a peaks were not always the local maximum of the a_p spectrum and a_p was smoothed by the instrument. On the other hand, the spectral decomposition of *Chase et al.* (2013) was spectrally unsmoothing a_p gaining in amplitude at the local $a_\varphi(676)$ peak. The a_{NAP} spectrum retrieved with *Chase et al.* (2013) was close to zero in a number of samples, this is likely due to the fact that the method focused on retrieving individual phytoplankton pigments and not a_{NAP}. A middle ground was found with the results from the method of *Zheng and Stramski* (2013), the stacked-constraints approach to partition a_p into a_φ and a_{NAP} from the latter was more robust and consistent through our dataset. From these observations, I recommend using the more advanced model of *Zheng and Stramski* (2013) to estimate a_φ and a_{NAP} from a_p.

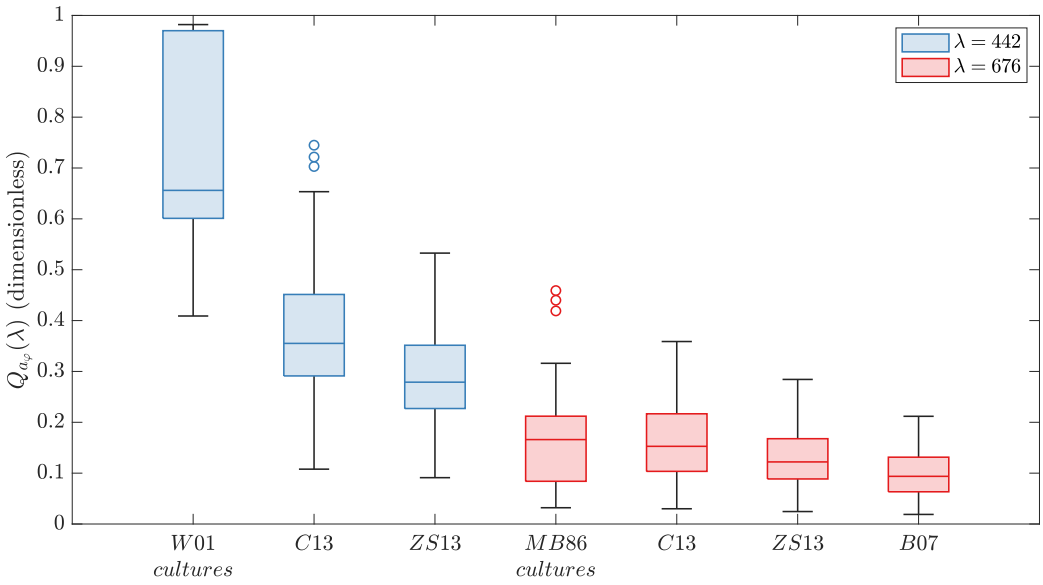

Figure 3.1. Phytoplankton efficiency factors at 442 nm and 676 nm for the NAAMES dataset computed with three different methods and compared to measurements reported in the literature. The culture measurements are Whitmire et al. 2010 (W01) and Morel and Bricaud 1986 (MB86). The methods used to retrieve the phytoplankton absorption were: Chase et al. 2013 (C13), Zheng and Stramski 2013 (ZS13), and Boss et al. 2007 (B07). On each box plot, the central mark indicates the median, the top and bottom edges indicate the 25^{th} and 75^{th} percentile. The whiskers extend to the most extreme data points excluding outliers (denoted by the 'o' symbol).

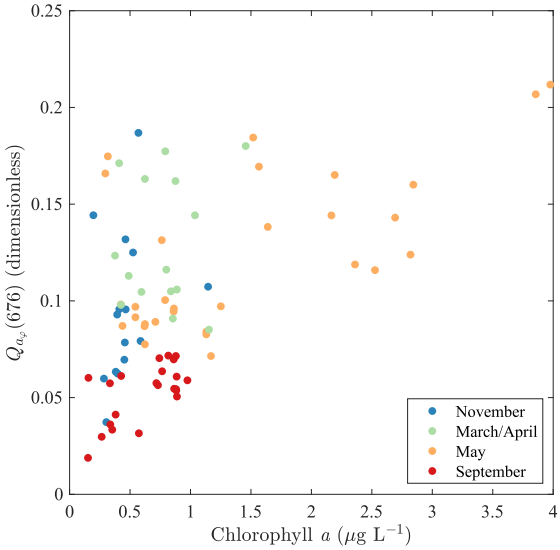

Figure 3.2. Phytoplankton absorption efficiency as function of biomass (chlorophyll a) and time of year (seasons). The low phytoplankton absorption efficiencies of September are consistent with less chlorophyll per cell which is expected for nutrient starved cell at high light levels.

3.3.2 Phytoplankton Attenuation

The contribution of phytoplankton to particulate attenuation in-situ was quantified in the western North Atlantic over multiple seasons. The ratio of phytoplankton to particulate attenuation, presented in Figure 3.3, were higher (73 % ± 23 % at 676 nm) than previously reported. Previous studies based on cytometry (no imaging) estimated that phytoplankton accounted for 50 to 60 % of particulate attenuation (*Durand and Olson*, 1996; *Green et al.*, 2003). While ratio of phytoplankton carbon (Cphyto) to particulate organic carbon (POC) had high variability across the Atlantic 12 % to 97 % with an average of 44 % (*Graff et al.*, 2015).

The comparison between the phytoplankton and total particulate size distribution (Figure 3.4) gave insights regarding why the phytoplankton dominates the attenuation signal in the NAAMES dataset. The particulate size distribution was dominated by phytoplankton between 0.6 µm and 66 µm ($G_\varphi/G_p = 76$ %± 23 %). However, at both extremities of the size distributions NAP were dominating; below 0.6 µm no phytoplankton is known to exist (also below our observation range) while above 66 µm the number of phytoplankton observed was not significant ($G_\varphi/G_p = 30$ %± 13 % in the size range 0.2 to 110 µm).

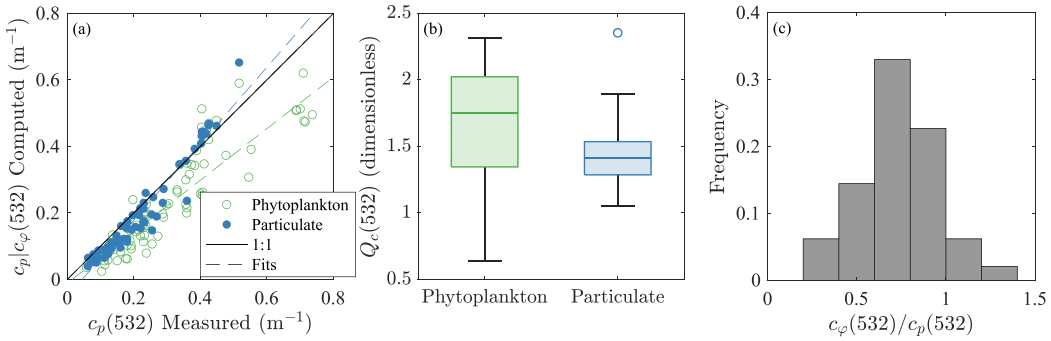

Figure 3.3. Phytoplankton (green) and particulate (blue) attenuation computed from their respective size distribution and compared to the measured particulate at 532 nm (a). (b) Attenuation efficiency factor of phytoplankton and particulate. (c) Contribution of phytoplankton to particulate backscattering.

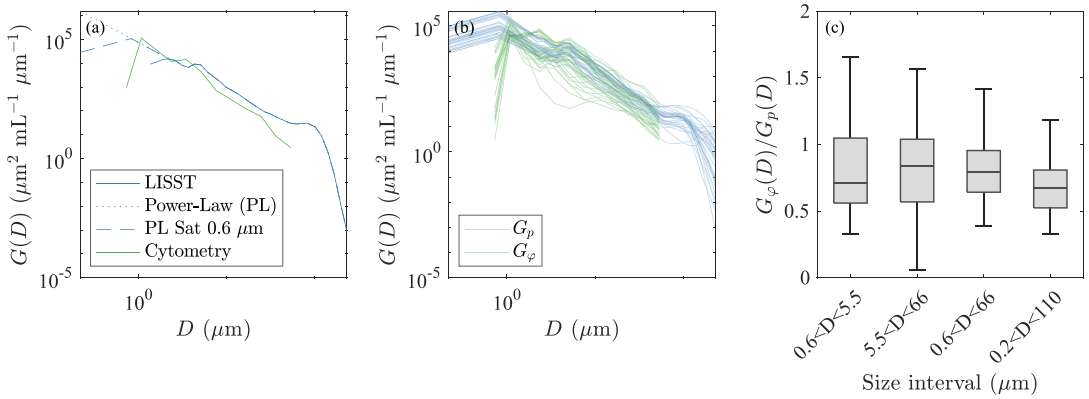

Figure 3.4. Phytoplankton and particulate area size distribution. (a) Particulate size distribution from the LISST (solid blue line) was extrapolated with a power law (dotted line) and a power-law saturating at 0.6 µm (dashed line). Phytoplankton size distribution was obtained from merged cytometry (green). A specific size distribution from the 4$^{\text{th}}$ NAAMES campaign is shown in (a) while the entire dataset (cytometry: all NAAMES campaigns and LISST: two most recent NAAMES campaigns) is shown in (b). The contribution of phytoplankton to the particulate cross-sectional area is presented for different size fractions in (c).

Phytoplankton attenuation efficiency factors ($Q_{c_\varphi}(532)$, Figure 3.3.b) was estimated to be 1.75 ± 0.23 at 532 nm for an index of refraction of $1.06 + i0.003$. Note that Q_{c_φ} was sensitive to both the wavelength (±0.14) and the index of refraction assumed (±0.25). The high abundances of pico-phytoplankton (mainly Synechococcus) contributed to an efficiency below two. Similarly, the extension of the particulate size distribution down to 0.2 µm size to which the particulate inherent optical properties were observed, significantly lowered the particulate attenuation efficiency factor (Q_{c_p}=1.41 at 532 nm). The particulate efficiency factor (Q_{c_p}) was very sensitive to the shape of the particulate size distribution between 0.2 µm and 2 µm. The sensitivity analysis of Q_{c_p} to the shape of the particulate size distribution (Figure 3.5) revealed that assumption made to extend the particulate size distribution of the LISST beyond 2 µm was reasonable with respect to optical modeling.

The acceptance angle of the instruments used were not taken into account when comparing to Mie theory. As the ACS and LISST had different acceptance angle (0.93° and

0.0269°, respectively) and their beam attenuation at 670 nm agreed well (bias <3 %, c.f. *Boss et al.* (2018b) for figure) I did not expect significant difference with the Mie calculation as the majority of the particles (>94 % of cross-sectional area on average) were smaller than 20 µm (*Boss et al.*, 2009).

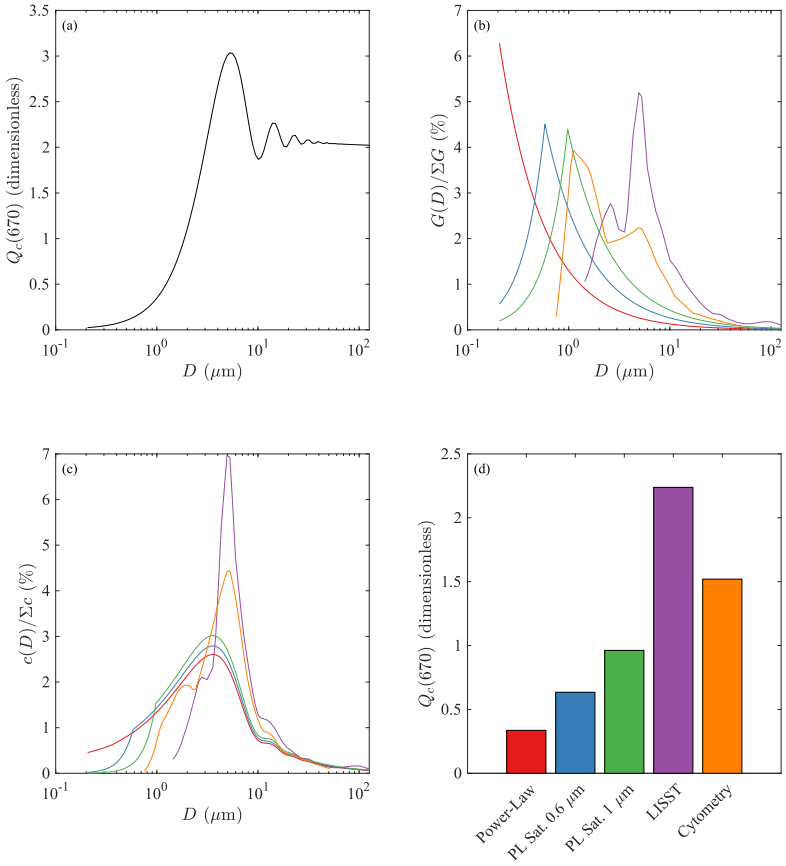

Figure 3.5. Sensitivity analysis of Q_c to the shape of the size distribution. (a) Attenuation efficiency factor (Q_c) at 670 nm as function of particle size (D). (b) The cross-sectional area (G) per size bin normalized by the total cross-sectional area for different shape of size distribution. (c) Contribution of each size bin to the total attenuation at 670 nm. (d) Efficiency factor computed for each size distribution. Size distributions are colored as follow: red: hypothetic power-law distribution of minus four; blue: same power law distribution saturated at 0.6µm; green: same power law size distribution saturated at 1.0 µm; orange: median NAAMES phytoplankton size distribution; purple: median NAAMES particulate size distribution observed by the LISST (all size bins).

The phytoplankton size distribution used in the present study assumed that all fluorescing particles (excited/emitted at 488/692 for small cells and 635/680 for larger cells) were phytoplankton. When looking at the images from the cytometer sizing larger cells (IFCB), 15 % of the cross-sectional area of all the images was classified as non-living (e.g. detritus, plastic) or zooplankton. The separation between phytoplankton and other fluorescing particles cannot be done with the non-imaging cytometer used to size the smaller cells. Hence, to be consistent in the merging of both the choice was made to keep this fraction of non-phytoplankton particles. Assuming this proportion of non-phytoplankton holds through the size spectrum then the total phytoplankton cross-sectional area would be overestimated by 15 % on average. This would lower the contribution of phytoplankton to IOPs by a similar amount.

3.3.3 Phytoplankton Backscattering

The particulate backscattering computed with the coated sphere model of *Zhang et al.* (1998) and the parameters from *Organelli et al.* (2018) was found to be in reasonable agreement below 0.2×10^{-3} m^{-1}, the range used to parametrize the model (Figure 3.6.a). For higher backscattering values the modelled particulate backscattering tended to overestimate backscattering (factor of 1.9). The backscattering efficiency factors at 532 nm retrieved from the coated sphere model with the size distribution were significantly lower than the one reported in *Organelli et al.* (2018) (0.011 +/- 0.002 for particulate and 0.017 +/- 0.004 for phytoplankton). The discrepancy could partially come from submicron particles (< 0.59 µm) as *Organelli et al.* (2018) did not integrate these particles. However, as our phytoplankton size distribution stops at 0.6 µm and equals or exceeds the measured particulate backscattering, it suggests that the current model or its parameters are inappropriate for our dataset.

The present modelling effort suggests that phytoplankton dominates the particulate backscattering signal (0.77 ± 0.21, Figure 3.6.c). This finding is similar to attenuation and

is consistent with the high ratio of phytoplankton to particulate cross-sectional area observed in the previous section. Nonetheless, the contribution of phytoplankton to particulate backscattering should be considered with caution as the model overestimated backscattering in our dataset.

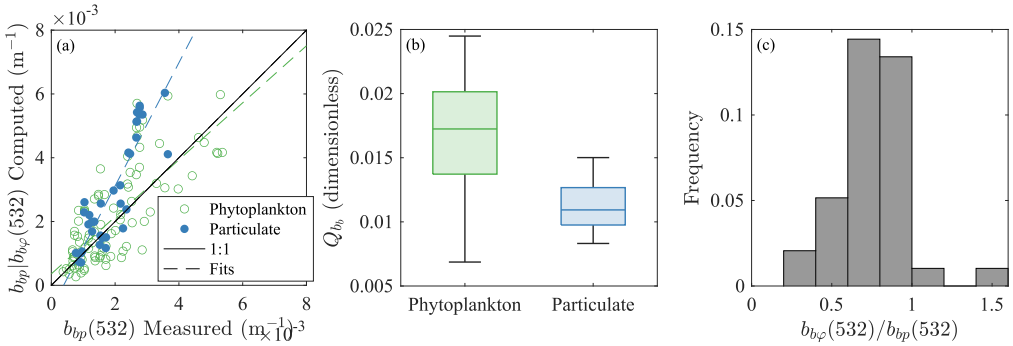

Figure 3.6. Phytoplankton (green) and particulate (blue) backscattering computed from their respective size distribution and compared to the measured particulate backscattering at 532 nm (a). (b) Backscattering efficiency factor of phytoplankton and particles. (c) Contribution of phytoplankton to particulate backscattering. The combination of parameters from Organelli et al. (2018) with a coat thikness (tk_2) of 7 % was used here.

3.4 Conclusion

To my knowledge, the efficiency of phytoplankton to absorb light (Q_{a_φ}), was measured for the first time in-situ across multiple seasons in the western North Atlantic. Q_{a_φ} values at 676 nm ranged from 0.1 to 0.2 (unit-less) which was similar to previous reported laboratory observations.

The comparison of particulate and phytoplankton size distribution showed that phytoplankton dominate (76 %) the cross-sectional area of the pull of particles in the open ocean across a wide range of physical and biological conditions. Hence, it reinforced the idea that particulate attenuation and backscattering are valid predictors of phytoplankton biomass and other parameters such as size. In fact, I found that phytoplankton contributed on average to 73 % of the total particulate attenuation.

The present modeling of particulate backscattering suggested that more research is needed to reproduce the backscattering measured in the open ocean across a broad range of environmental conditions. While the backscattering modeling of this study, using coated-spheres, was in agreement with measurements and modelling of *Organelli et al.* (2018) at low particulate concentrations ($b_{bp} < 2 \times 10^{-3}$ m^{-1}), it overestimated particulate backscattering at higher concentrations. The present study provides a unique dataset to test novel optical models and points out the need for an improved characterization of the size distribution of submicron particles.

CHAPTER 4
CONTRIBUTION OF ZOOPLANKTON AND SMALL MESOPELAGIC ORGANISMS TO BACKSCATTERING AND ULTRA-VIOLET FLUORESCENCE PROFILES

This chapter has previously been published in *Geophysical Research Letters*: Haëntjens, N., Della Penna, A., Briggs, N., Karp-Boss, L., Gaube, P., Claustre, H., & Boss, E. (2020). Detecting mesopelagic organisms using biogeochemical-Argo floats. Geophysical Research Letters, 47, e2019GL086088. https://doi.org/10.1029/2019GL086088

4.1 Introduction

Diel vertical migration (DVM) of mesopelagic organisms plays an important role in the biological pump by exporting particulate organic carbon (POC) from the euphotic layer of the ocean to the twilight zone (*Steinberg and Landry*, 2017). Traditional methods to sample the vertically migrating animals include sampling with nets and via acoustic remote sensing. However, given the limitation of sampling by research vessel (limited spatial and temporal coverage, avoidance of the nets from the fastest organisms, lack of taxonomical resolution from acoustic sensors), there is a strong interest in complementing such sampling with gliders (*Ohman et al.*, 2019; *Benoit-Bird et al.*, 2018) and profiling floats if they could provide information on the organisms' distribution (*Boyd et al.*, 2019).

The array of biogeochemical-Argo (BGC-Argo) profiling floats (*Claustre et al.*, 2020) measures water's physical, chemical (e.g. pH, oxygen, and nitrate), and biological properties (e.g. chlorophyll *a* fluorescence and backscattering). Methods have been developed to help constrain estimates of carbon uptake, and several particulate carbon flux pathways have been successfully estimated using bio-optical sensors mounted on floats (*Dall'Olmo and Mork*, 2014; *Llort et al.*, 2018; *Estapa et al.*, 2019). The premise of some of these approaches is to attribute spikes in backscattering and chlorophyll *a* fluorescence,

from sensors mounted on gliders, to sinking aggregates (*Briggs et al.*, 2011). To date, signatures of migrating mesopelagic organisms in float profiles have only been briefly mentioned but not thoroughly investigated in the literature *Boyd et al.* (2019). Previous studies have suggested that periods of spikes in optical profiles were associated with the periodic interaction of zooplankton with transmissometers (*Bishop et al.*, 1999; *Bishop and Wood*, 2008). Zooplankton are also associated with backscattering "spikes", defined here as a sharp and large increase in the optical signal (similar to *Briggs et al.* (2011) and detailed in Supporting Information), in measurements collected by sensors mounted on a surface underway system and on an underwater stationary observatory (*Burt and Tortell*, 2018; *Tanaka et al.*, 2019). These observations suggest that the presence of a subset of mesopelagic organisms could be estimated using BGC-Argo optical profiles and that the presence of spikes in optical instruments should not be interpreted a priori as an indication of sinking particles.

Here we combine optical profiles from floats visited or deployed during the North Atlantic Aerosols and Marine Ecosystems Study (NAAMES, *Behrenfeld et al.* (2019)), with nearby observations in the vicinity from scientific echo sounders, an Underwater Vision Profiler (UVP), and a mesopelagic net tow to determine associations between spike layers and the vertical distribution of mesopelagic organisms. We then propose an algorithm to detect the presence of mesopelagic organisms that can be applied in other float studies. The algorithm is applied to the entire BGC-Argo network, from which diel and seasonal distribution pattern of spike layers emerge and are compared to the literature on the migrations of mesopelagic organisms.

4.2 Materials and Methods

4.2.1 Field Measurements

During the NAAMES campaigns, which included four cruises on board the R/V Atlantis over four different seasons (November 2015, March-April 2018, May 2016, and

September 2017), 15 BGC-Argo floats were deployed (13 Sea-Bird Navis and 2 NKE PROVORS). Ten of them were revisited during following campaigns (*Behrenfeld et al.*, 2019, Table 1). Float profiles and ship-based measurements were defined as "concurrent" when float profiles were at a distance not exceeding 50 km away from the ship and within eight days following the time the ship occupied a station. One float was excluded from the analysis due to malfunction of one of its optical sensors. Altogether, we obtained 20 float-ship matchups with 2 to 24 profiles per matchup, an average distance between the ship and the float of 22 km, and an average time difference between the ships station and the float profile of 1.4 days (Table B.1). At most stations, acoustic measurements, CTD casts with a UVP, and a mesopelagic net tow were carried out.

Most NAAMES floats, including the one from the case study presented below, were Sea-Bird Navis BGCi equipped with a CTD (Sea-Bird SBE 41N), dissolved oxygen sensor (Sea-Bird SBE 63), a photosynthetically available radiation (PAR) sensor (Satlantic OCR 504), and a bio-optical sensor (WetLabs ECO-MCOMS) to measure chlorophyll *a* fluorescence (fchl), backscattering at 700 nm (b_b), and FDOM (fluorescence by colored dissolved organic matter, CDOM). The excitation/emission wavelengths of the optical sensor's LEDs are centered at 370/460, 470/695, and 700 nm for FDOM, fchl, and b_b respectively. Only measurements taken during the float's ascent to the surface are recorded. The floats' optical sensors readings were converted to scientific units using the manufacturer's coefficients according to the standard BGC-Argo protocols (*Schmechtig et al.*, 2015, 2017, 2018). We further calibrated fchl with High Performance Liquid Chromatography (HPLC)-derived chlorophyll *a* collected with CTD profiles near the float ($chla_{HPLC} = (chla_{manufacturer} - 0.019)/2.32$), following the recommendation of *Roesler et al.* (2017). Depth was computed from the measured pressure using TEOS-10 (*McDougall*, 2011) to compare with the acoustics observations.

A scientific echo sounder (BioSonics DT-X) was used to continuously measure acoustic backscattering with two single-beam transducers transmitting at 38 kHz and 120 kHz.

Acoustic volume scattering strength, S_v, was computed using a custom written Matlab software that incorporates a time varied gain (TVG) correction that takes into account the loss in transmission due to acoustic beam spreading and the absorption by the acoustic medium. We followed the method described in *De Robertis and Higginbottom* (2007) to quantify and limit the effect of background noise and the approaches described in *Ryan et al.* (2015) to filter out the majority of transient and impulsive noise, as well as attenuation due to surface bubbles. The top 10 meters of the acoustic data were removed as bubbles and near-field effects contaminated the observations there.

A strong thermocline was observed at the sampling stations (18.3 °C at 24 m, and 6.5 °C at 67 m, Figure 4.2). To account for changes in sound speed across this density interface, we computed both the acoustic attenuation coefficient (α) and the sound velocity (c) using the models of *Francois and Garrison* (1982) and *Medwin* (1975). The constant sound velocity was converted into time-varying sound speed profiles to estimate the corrected depth of each backscattering bin. The corrected depth was on average 5 m shallower at 330 m and 10 m shallower at 545 m than the depth obtained assuming constant T and S. The uncertainty of the corrected depth is <1 m considering instrument resolution, error in temperature measurements (<1 °C) and shifts in the thermocline (<10 m). The depth of the scattering layer observed could however be affected by the presence of the ship. Surface scattering layers at night were observed 25 to 75 m deeper than in the absence of ship in previous studies (*Ona et al.*, 2007; *De Robertis and Handegard*, 2013; *De Robertis et al.*, 2019; *Peña*, 2019).

The size of particles matching the wavelength of the sound emitted by the two transducers (and hence likely to have the largest acoustic backscattering return per mass) is on the order of 1 and 4 cm for 120 and 38 kHz, respectively.

The ship's CTD Package was equipped with a beam transmissometer, a flat-faced chlorophyll *a* fluorometer and backscattering sensor (WetLabs ECO-FLNTU), and an Underwater Vision Profiler (Hydroptics UVP 5HD). CTD profiles were processed with

Sea-Bird software and calibrated with manufacturers coefficients. The UVP was used to image particles greater than 600 µm (mostly zooplankton and marine snow, *Picheral et al.* (2010)), and provided information on particle size distribution for particles larger than 64 µm. Images and their associated metadata were imported to EcoTaxa, a web based platform for image curation and classification (*Picheral et al.*, 2017). EcoTaxa's algorithm to predict image classes used features (e.g. Feret diameter, area, equivalent spherical diameter) extracted from the images using a custom ImageJ-based software (Zooprocess *Abràmoff et al.*, 2004). Due to the shape and the descent speed of the CTD package and relatively low abundance of large zooplankton in the UVP's field of view, the abundance and distribution of macro-zooplankton was not well resolved by the UVP (*Picheral et al.*, 2010); several profiles were aggregated to obtain sufficient number of particles (when available).

Isaacs-Kidd Midwater Trawls (IKMT) were conducted at each station to provide a qualitative description of the mesozooplankton and mesopelagic fish community. The net was deployed during the evening for 40 minutes at 300-350 m and towed at a speed of ∼1.5 knots. The collected organisms were classified, preserved, and stored. This type of net tow is valuable to describe the bio-diversity of mesopelagic fish. The mesh size (3 mm) of the net is too large to properly quantify the abundance of smaller specimens (e.g. *Crustacea*, *Decapoda*, *Euphausia*).

4.2.2 Analysis of BGC-Argo Database

To assess spatial distributions of optical spike layers on a global scale, the analysis of the NAAMES floats was augmented with an analysis of all BGC-Argo synthetic profiles (*Bittig et al.*, 2018) that were equipped with backscattering sensors and were available at time of writing this manuscript (September 17, 2019). Considering the extensive set of profiles (n=70246 BGC-Argo profiles), an algorithm was developed for automatic identification of layers of spikes. These layers are characterized by a high density of spikes

in a short vertical range, typically 15 spikes within 50 m. This approach effectively rules out floats that have low sampling resolution (<1 observation/2 dBar, Supporting Information). The spike layer detection algorithm first finds individual spikes in the profile and then identifies if the spikes cluster into layers. Details of the implementation as well as the assessment of its performance are presented in the Supporting Information. For the analysis of layer of FDOM and b_{bp} spikes presented here, each layer of spikes detected by the algorithm was manually validated. Thus, the distribution of spike layers can only be underestimated as the algorithm might have missed some profiles with spike layers (14 % for FDOM and 35 % for b_{bp}, Table A.1). Information extracted from the spikes of fchl are not presented as their origin remains unclear (Section 4.3.3) and the performance of the spike layer detection algorithm are not reliable with this channel (Supporting Information). To further study the relationship between the occurrence of spikes and the diel vertical migration of mesopelagic organisms the sun elevation was computed for each profile using *Reda and Andreas* (2004) as a PAR sensor was not available on all floats. Comparison of the relationship between layer depth and sun angle or PAR for floats who had such sensors yielded similar patterns (not shown).

4.3 Results and Discussion

4.3.1 NAAMES Observations

To illustrate the hypothesis that mesopelagic organisms are at the origin of spike layers in float profiles, we showcase one of the three stations for which we have collocated float data and ship observations. At the case study station (Figure 4.1), acoustic backscattering at 38 kHz reveals the timing and depth variability of the diel vertical migration (DVM) of organisms from ∼50 m at night to ∼350 m during daytime. This location is also characterized by a non-migratory deep scattering layer (DSL) observed between 400 and 550 m. Observations collected by the 120 kHz transducer, which are better suited for detection of smaller organism but limited to the upper 200 m of the water column, also

show the migrating layer with returns magnitudes slightly higher than those at 38 kHz (1.7 ± 5.7 dB). The location of deep and surface scattering layers detected with the UVP data is consistent with the ones from the acoustic backscatter: during night-time, higher abundances of zooplankton are observed in the upper 200 m with a peak at 50 m matching the acoustic migrating layer, while during daytime higher abundances of zooplankton are observed at depth (Figure 4.2.c). Layers of FDOM and b_{bp} spikes were detected in float profiles during daytime (including dawn and dusk) at depths slightly shallower than the ones of the DSL observed with the acoustic (Figure 4.1) and the UVP (Figure 4.2).

From the concurrent observations of bio-optical float profiles and ship-based acoustics of the NAAMES expeditions, three matchups had spike layers in float FDOM profiles (15 %) and five matchups had individual spikes in float FDOM profiles (25 %, Supporting Information). The analysis revealed that layers of FDOM and b_{bp} spikes were always in the presence of a strong DVM but that not all strong DVM were associated with layers of spikes. Two patterns emerge from our limited dataset: (1) the depth of the spike layers or individual spikes from floats is correlated with the depth of the diel migrating acoustics layer while; (2) the depth of the spike layers is mismatched from the depth of the acoustically measured scattering layer, typically at a shallower depth (on average 42 m above, but up to 80 m).

Several hypothesis could be brought forward to explain these observations. The depth offset could originate from our observational method due to: (1) distance between ship and float, (2) difference in sampling time, (3) deepening of the acoustic layer due to the ship's presence, and/or (4) the float is observing a subset of the migrating mesopelagic population. The first two hypothesis could explain why the depths are not consistent but not the presence of a consistent bias. The third hypothesis is unlikely as, to our knowledge, no deepening of scattering layers at such depths was reported (*Ona et al.*, 2007; *De Robertis and Handegard*, 2013; *De Robertis et al.*, 2019; *Peña*, 2019). The fourth hypothesis cannot be rejected and is is discussed further below.

Multiple studies have shown that organisms' size, species richness, and community composition vary with depth: distinct species assemblages are present at different depth (*Sommer et al.*, 2017; *Kosobokova et al.*, 2011; *Kosobokova and Hopcroft*, 2010; *Steinberg et al.*, 2008a,b; *Vinogradov*, 1970). NAAMES echograms partially illustrate this community's stratification; for example, in Figure 4.1 four layers of varying intensities can be observed during the daytime: two very thin layers at 100 and 150-200 m respectively and two wider layers at 300 to 400 m and 400 to 550 m. None of the observational methods captures the full size range of mesopelagic organisms, rather each method is better suited for a specific size range and type of organisms (*Vinogradov*, 1997). The acoustic frequencies used here are more sensitive to larger organisms (13 and 39 mm) while the UVP is designed for organisms larger than 0.6 mm (*Picheral et al.*, 2010), suggesting that the depth mismatch between the float and the ship-based observations could result from the stratification of communities of organisms with depth. As the organisms observed at the origin of the spike layer are shallower than the acoustics scattering layer, they are likely to be smaller. *Ohman and Romagnan* (2016), report that the vertical distribution shallows progressively with decreasing copepod body size. Thus, the spike layer observed by the float are likely caused by a narrow subset of mesopelagic organisms not prominent in the UVP images and the acoustics scattering.

4.3.2 Light Attraction of Mesopelagic Organisms

When floats are profiling above 1000 m, the LED (UV, blue, and red) of the bio-optical sensors are continuously blinking and potentially attracting mesopelagic organisms. Positive phototaxis is well known for mesopelagic organisms, in fact, light traps are commonly used to sample marine biodiversity at depth (*McLeod and Costello*, 2017). This positive phototaxis has also been observed with similar bio-optical sensors in coastal environments (*Tanaka et al.*, 2019).

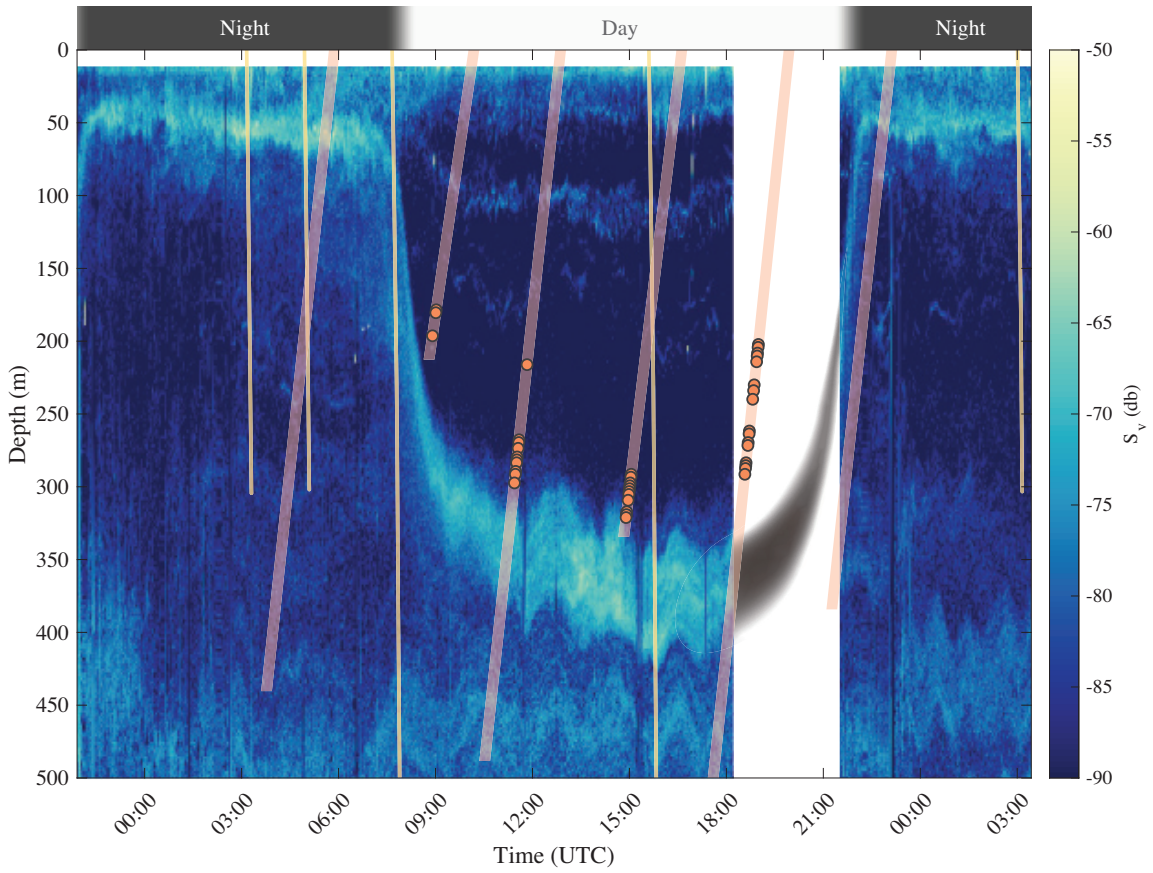

Figure 4.1. Time-series of the mean volume backscattering strength (S_v) at 38kHz measured from the pole-mounted echo sounder at the NAAMES station located at 44° 21.838'N, 43° 21.503'W occupied from September 5, 2017 21:00 to September 7, 2017 3:00 UTC. Another acoustic frequency (120 kHz) and other stations with float-acoustics matchups are presented in the Supporting Information. The yellow lines correspond to the down-cast of the CTD and the UVP deployed from the ship. The orange lines are the up-casts of the float and orange circles superimposed on the profiles are FDOM spikes. Note that the first, third and fifth orange lines are profiles of the floats from the following day. The data from these float profiles are presented in Figure 4.2. The slopes of the lines correspond to the profiling speed (0.08 m/s for the float and 0.5 m/s for the CTD). No acoustic backscatter was collected between 18:12 and 21:29, the gray shape indicates an estimated depth of the main DVM scattering layer, the timing was estimated from the DVM observed at other NAAMES stations. The first 10 m of the data are removed to mask near-field effects in the acoustics signal.

If mesopelagic organisms are attracted to the floats, are they able to reach the light source? When considering the average ascent rate of floats (∼0.08 m/s), it is plausible to assume that some species of zooplankton and other mesopelagic fish are capable of keeping

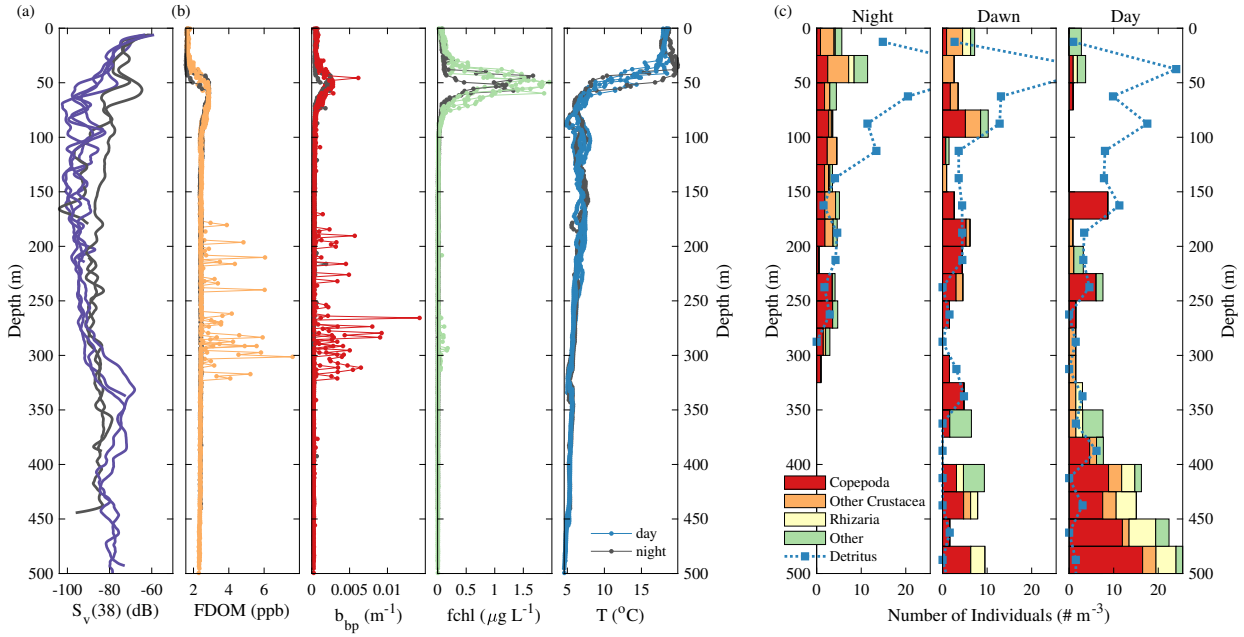

Figure 4.2. Profiles of the float, the acoustics, and the UVP at the same station presented in Figure 4.1. *(a)* Profiles of acoustic backscattering at 38 kHz ($S_v(38)$) from the ship. *(b)* Profiles of fluorescent dissolved organic matter (FDOM), particulate backscattering (b_{bp}), chlorophyll *a* fluorescence (fchl), and temperature (from left to right) from the float (WMO 5903108) deployed upon arrival on station showcased (Figure 4.1). Black lines represent to night-time profiles while colored lines (purple, orange, red, green, and blue) correspond to day-time profiles. The time of the acoustic profiles match exactly with the time of the float profile. *c)* Profiles of *copepoda* (red), other *crustacea* (orange), *rhizaria* (yellow), other zooplankton (*Annelida*, *Chaetognatha*, *Cnidaria*, and unidentifiable; green) and detritus (blue lines) binned by 25 m observed by the UVP deployed from the ship. The night profile (left) is an average of the 1st, 2nd, and 5th casts on station, the dawn (middle) and day (right) profiles are based on a single cast. Some detritus concentrations are out of scale (night cast 41 particles/m^3 and dawn cast 35 particles/m^3).

up with floats. Estimated DVM speeds of mesopelagic organisms based on the acoustic data are 0.06 ± 0.01 m/s upward and downward (consistent with *Bianchi and Mislan* (2016)) and the observed active swimming speeds of *Crustaceans* and micronekton range from 0.02 to 0.10 m/s and from 0.025 to 0.30 m/s respectively (*Ignatyev*, 1997).

In contrast to the float profiles, no spike layers were observed in the particulate attenuation (c_p), b_{bp}, and fchl deployed on the CTD Rosette deployed at the same station. A few individual spikes were observed in c_p and b_{bp} at 150−250 m and 75 m respectively at

∼3:00 UTC but their depths was not associated with DVM scattering layer, and no spikes were observed in the fchl profile from the CTD Rosette. The CTD Rosette and associated sensors with its 24 12 L bottles displaces a considerable volume of water while travelling down and up at speeds significantly faster than floats (0.5-1 m/s). This likely prevents mesopelagic organisms from being able to reach the sensors if attracted to them, especially with the strong shear surrounding the CTD package. These results suggest that the floats are not measuring ambient animal concentration but rather that of mesopelagic organisms attracted to the lights of the bio-optical sensors.

Furthermore, a comparison of the pairs of upcast and downcast of the PROVOR floats (only platform for which downcast data can be recorded) in the western North Atlantic (n=31) indicates that downcasts tend to have twice as many spikes as upcasts. Given that the PROVOR downcasts are significantly slower than the upcasts (0.04 m/s slower on average) with the same sampling frequency suggest that the elements at the origin of the spikes are staying in front of the sensor, which further supports the light attraction hypothesis. However, this comparison also showed that for up and down casts performed within 2 hours (n=8) the depth of the layers was not significantly different, therefore that the phototactic behavior is not sufficient to explain the depth mismatch between the floats and the acoustics.

4.3.3 The Origin of the Spikes

While the present results strongly suggest that layers of FDOM or b_{bp} spikes are associated with DVM mesopelagic organisms, the stability of the sensor was assessed to rule out any electronic noise. We then attempt to narrow the variety of organisms at the origin of the spikes.

The sensors used with the BGC-Argo floats (WetLabs ECO-MCOMS or ECO-FLBBCD) have been extensively characterized (*Carroll et al.*, 2005; *Twardowski et al.*, 2007; *Cetinić et al.*, 2009; *Sullivan et al.*, 2013) and have been found to be

susceptible to power surge, submersion, and electromagnetic fields (M. Slivkoff personal communication). Since the majority of profiles from many different platforms does not contain any FDOM, b_{bp}, or fchl spikes and there is no evidence of the phenomenon listed above, we excluded sensors and platform effects as a source of the spikes observed. We then tested if spikes of similar intensities could be produced by zooplankton using sensors similar to the one mounted on the float (Supporting Information). Like *Tanaka et al.* (2019), who used cameras looking at an fchl and b_{bp} sensor, we confirmed that spikes in FDOM, b_{bp} and fchl could originate from zooplankton. Spikes in backscattering have long been associated with the presence of large particles. Spikes in FDOM can be explained by the fact that fluorescent proteins are possessed by a range of marine organisms and that both humic- and amino acid-like fluorescence are produced by zooplankton grazing and excretion (*Stedmon and Cory*, 2014). Spikes observed in fchl co-occurring occasionally with the layers of spikes observed in FDOM and b_{bp} channels are harder to explain. A possible explanation could be the fluorescence from the gut of the zooplankton. However, it would likely be observed only for a short period (several hours) after the time of grazing which isn't the case in our dataset. Spikes in fchl could also be caused by the same fluorescent compounds seen by the FDOM sensor (*Xing et al.*, 2017) or by an out of band response given the near saturation of the backscattering channel when measuring spikes which uses the same detector but a different light source.

In order to identify the organisms that were associated with the observed spikes, we analyzed the community composition of zooplankton observed with the UVP, the size of the organisms observed by the acoustic, and the community composition of mesopelagic fish sampled with the IKMT net. During the case study station, the zooplankton population was highly dominated by *Copepoda* and *Euphausiids* (Figure 4.2.c) with most living organisms seen by the UVP having a Ferret diameter >5 mm and <25 mm. The acoustics used to observe the DVM layer was mainly sensitive to organisms with a length of 1 and 4 cm. The samples of the IKMT net, targeting the DSL layer and therefore containing both

DVM and non-DVM organisms, were dominated by *Myctophids* (*Diogenichthys atlanticus*, *Benthosema glaciale*, and *Myctophym punctatum*) during stations of NAAMES 3 with spike layers, and by *Bristlemouths* (*Cyclothone pseudopallida*) and *Penaeidae* during the stations of NAAMES 2 with spike layers. These organisms had lengths ranging from 2 to 11 cm. While we could characterize the larger dominant taxa present in the water column, we could not identify the subset of smaller organisms likely at the origin of the spike layers.

4.3.4 BGC-Argo Observations

Analyses of FDOM and b_{bp} spikes from the broader BGC-Argo database (Figure 4.3) revealed the following patterns: (1) layers of spikes were mainly observed at greater depth during daytime while closer to the surface during nighttime, (2) layers of spikes are more abundant from the end of spring to early fall in North Atlantic and similar seasons in the Southern Ocean, and (3) layers of spikes are observed primarily in North Atlantic and secondarily in the Southern Ocean. The observations in the North Atlantic are consistent with the current understanding of zooplankton life-cycle in this basin: enhanced activity during the phytoplankton spring bloom and summer followed by a reduced activity and deepening (diapause) during the winter (*Miller and Wheeler*, 2012; *Colebrook*, 1982; *Falkenhaug et al.*, 1997; *Heath*, 2000).

In addition, fewer profiles with a layer of spikes are observed at night than during the day (Figure 4.4). Perhaps, feeding, one of the primary reasons for DVM (*Hays*, 2003), could explain the reduction in the phototactic behavior of the taxa responsible for the spike layer. Few profiles are available in the tropics preventing any strong conclusions about this area, only 4 spike layers were observed in the tropics or sub-tropics. Still, the normalized (i.e. accounting for sampling effort) number of profiles with spikes is lowest in the tropics and sub-tropics compared to other parts of the ocean (0.16 % vs 2.6 %).

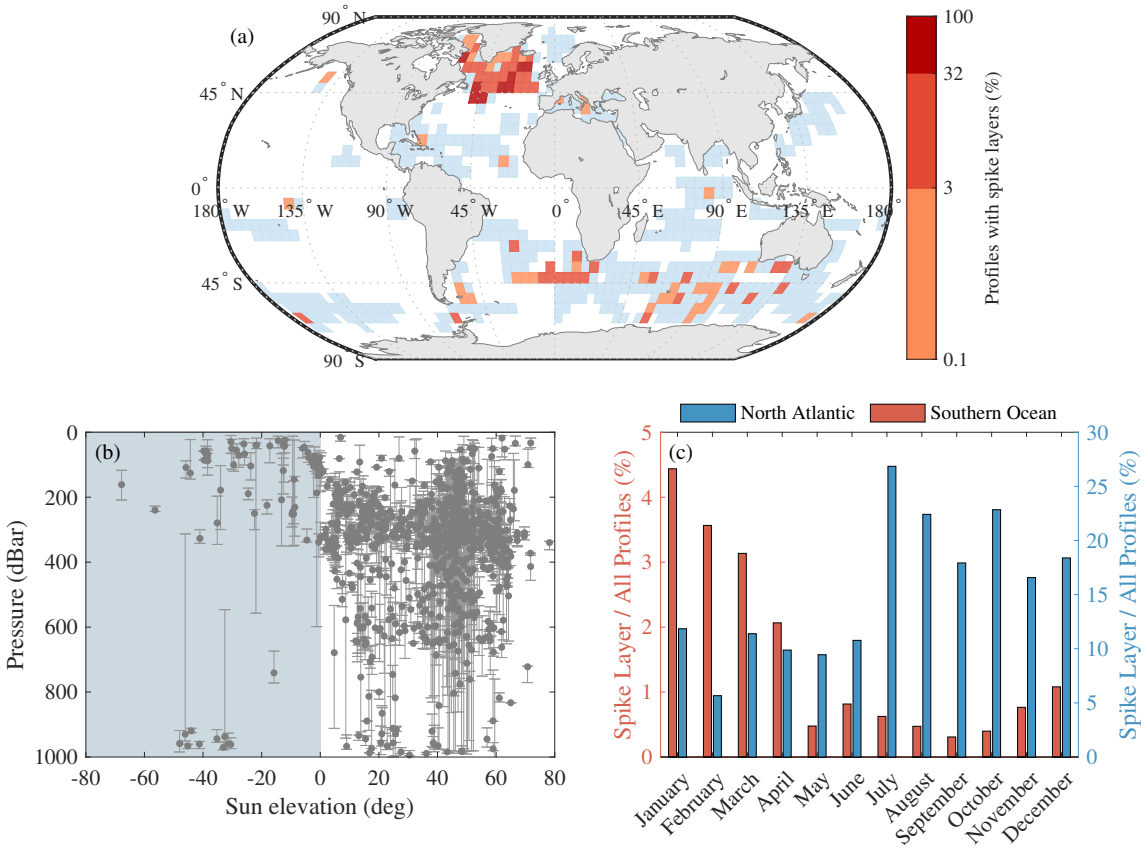

Figure 4.3. Analysis of BGC-Argo floats with a profiling resolution greater than 1 observation/3 dBar. *(a)* Locations of float profiles in a 5×5° grid, blue areas indicate the presence of float profiles and red areas indicate the percentage of profiles with b_{bp} spike layers. *(b)* Average depth of b_{bp} spike layer as function of the sun elevation, the error-bars correspond to the depth of the shallowest and the deepest spike of the layer. Replacing the sun elevation with PAR measured by the floats equipped with such sensors does not change the pattern significantly (not shown). *(c)* Number of profiles with b_{bp} spike layers normalized by the total number of profiles in a region per month, the blue histogram is for North Atlantic area while the red histogram correspond to the Southern Ocean. Additional information regarding the statistics of the spike layers can be found in the Supporting Information.

4.3.5 Other Considerations

We have applied the algorithm developed to automatically detect the presence of spike layers in bio-optical profiles. It performed well in identifying spike layers where layers were recognized visually (Table A.1) with FDOM and b_{bp} profiles but was not reliable for fchl profiles (Supporting Information and Section 4.3.3). The algorithm offers a way to resolve the source of the b_{bp} spikes; randomly distributed spikes of b_{bp} which are not accompanied

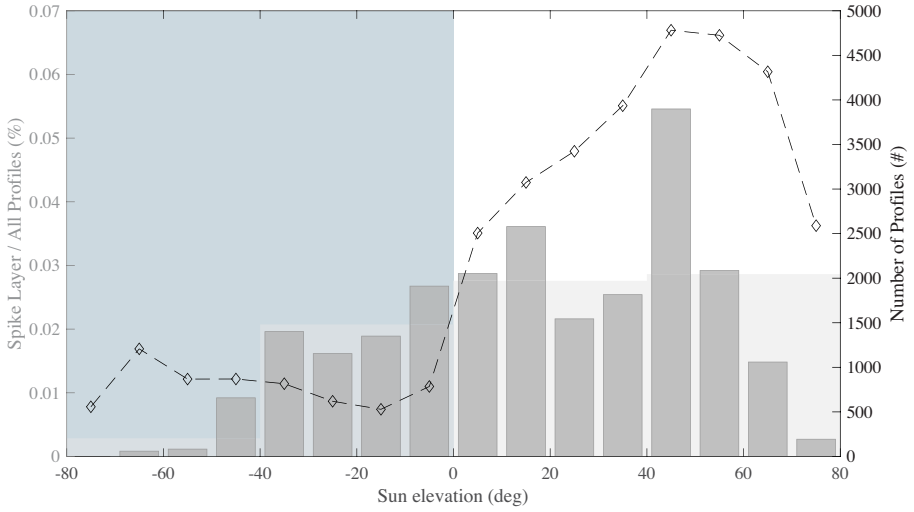

Figure 4.4. Number of spike layers events normalized by sampling effort (number of profiles) as a function of the sun elevation (zenith angle) (gray histogram). Dark grey histogram correspond to 10^o bins and lighter histograms in the background correspond to 40^o bins. The total number of float profiles as a function of sun elevation is overlaid as a dashed black line. Negative sun elevation correspond to night time while positive sun elevation correspond to daytime.

by FDOM spikes are likely due to sinking aggregates (*Briggs et al.*, 2011) while b_{bp} spikes grouped in layers accompanied by FDOM spike layers are likely due to mesopelagic organisms. Thus, mounting FDOM sensors on profiling floats can help distinguish sources of b_{bp} spikes in float profiles. Nonetheless, the method proposed on floats without FDOM channels works well (identifying correctly layers with a probability >94 %). We also recommend to visually inspect and validate each spike layers found.

The results in Figure 4.3 are likely biased by additional factors: (1) The excitation wavelength of the optical sensors affect the number of spike layer observed: floats with a backscattering channels at 532 nm (8093 profiles in the BGC-Argo database) exhibited 50 % more spike layers than the other floats in North Atlantic. (2) The profiling frequency: several PROVOR floats from the BGC-Argo array have been setup to profile on a diel basis at times during which the spike layers were prominant which significantly increased the number of spike layers observed. It also allows to observe DVM cycles directly from a float (*Boyd et al.*, 2019).

The frequency of spikes in optical signals is tightly linked to the sampling frequency of the sensors. *Briggs et al.* (2011) recommends having >200 sampling points within a 50 m interval for addressing single spike events due to aggregates. Here, the spikes layers of interest are likely due to phototactic organisms, therefore a lower sampling frequency seems sufficient. The positive phototaxis suggests that bio-optical sensors must be continuously sampling in order to detect spikes as organisms are less likely to swim toward a sensor flashing at low frequency (e.g. once every 30 s). The majority of spike layers are observed with sampling resolution >1 observation/2 dBa. On the other end, no spike layers could be identified on floats with low sampling resolution <1 observation/5 dBar. De facto, this has filtered out the use of APEX BGC-Argo floats (24873 profiles) from this analysis (see Supporting Information). The float WMO 5903108, presented in Figure 4.1 and 4.2, was set up to sample discretely from a 1000 dBar to 100 dBar with the bio-optical sensor continuously running (personal communication with Sea-Bird Scientific) resulting in 1-1.5 observations/dBar which was sufficient to detect the presence of mesopelagic organisms with 2 to 18 spikes/50 dBar. Based on this analysis we recommend that to consistently detect spike layers, floats should profile at a speed <0.08 m/s (i.e. mesopelagic organisms can keep up with the floats), with a sampling resolution >1 observation/dBar (e.g. the bio-optical sensor continuously sampling above 1000 m), and for the purpose of studying the DVM, floats should profile several times a day.

4.4 Conclusion

Multiple collocated observations during the NAAMES campaigns and from the BGC-Argo dataset show diel variations in the depth of the spike layers that are detected by bio-optical sensors. This pattern is consistent with diel vertical migration of zooplankton and mesopelagic fish and suggests that a subset of mesopelagic organisms are attracted to the light generated by the bio-optical sensors that are mounted on the floats. Their presence particularly impacted channels designed to measure FDOM and b_{bp}. While the

exact organism(s) that generate the spike layers is uncertain, concurrent observations from the UVP and IKMT suggest that *Copepoda*, *Euphausiids*, *Myctophids*, and/or *Bristlemouths* were present in the water column. As part of this analysis, a novel method was developed for the detection of spike layers in FDOM or b_{bp} profiles allowing to distinguish the types of particles at the origin of b_{bp} spikes (aggregates or zooplankton). This method can be applied to other slow profiling platforms with high resolution bio-optical sensors such as gliders or wire-walkers. For the first time, BGC-Argo floats can be used to obtain information on organisms other than phytoplankton, adding higher trophic levels to their sensing capability by providing an opportunity to confirm the presence of specific DVM organisms. Furthermore, this type of analysis provides a unique opportunity to study organisms not prominent in current acoustic datasets, with associated hydro-graphic variables. We encourage future work to identify the specific taxa at the origin of the spike layers. In the near future, floats equipped with cameras (Hydroptic, UVP6-LP) and active acoustics sensors (miniature sonar, *Goulet et al.*, 2019) provide a promising direction to further study these organisms.

CHAPTER 5

CONTINUOUS DATA LOGGING SOFTWARE FOR SHIPS UNDERWAY OPTICAL INSTRUMENTATION

This chapter has previously been published in *DIY Oceanography*: Haëntjens, N., and E. Boss. Inlinino: A modular software data logger for oceanography. Oceanography 33(1):80–84. https://doi.org/10.5670/oceanog.2020.112

5.1 Introduction

Our laboratory equips research vessels with optical instruments to continuously measure the properties of the ocean while underway, offering extended spatial and temporal coverage of the state of the surface ocean (*Boss et al.*, 2019). These observations are used, for example, to assess phytoplankton biomass and size (*Graff et al.*, 2015; *Boss et al.*, 2015) and suspended sediments (*Dierssen et al.*, 2006b). These products are computed from optical properties such as fluorescence, backscattering, absorption, and attenuation. Each instrument used (e.g. WET Labs ECO-BB3, WET Labs WETStar WSCD, WET Labs ACS) measures a specific property that needs to be recorded externally. Given the instrumentation involved and our needs, the data logging system has to be capable to (1) interface with instruments through serial (RS-232) or analog (0 to 5 V) communication, (2) accurately time stamp the data received with the research vessel NTP server, (3) store the data in files easily readable by data analysis software (e.g. R, MATLAB), and (4) be able to provide data visualization for real-time monitoring of the measurements.

We first looked at the instruments' manufacturer (WET Labs) logging solutions that may match the specifications above. They designed systems for two broad situations: (1) short time use in a laboratory environment (e.g. testing equipment or experiments of a few hours); (2) autonomous deployment underwater (e.g. profiling the ocean to several hundreds of meters). While both solutions worked, we found them impractical for

continuous month-long use. The first solution was generating oversized data files in our use case while the second solution was found to have a drifting timestamp when used over extended periods in addition to having memory limitation (e.g. limited to the memory of the device, hour-long data download from the data logger to computer preventing data acquisition).

We then looked for alternatives and came across many commercial hardware data loggers (e.g. Campbell Scientific). While hardware data loggers would have likely suited our requirements, we oriented our approach toward software data loggers running on computers as they presented the following advantages: computers could easily be synchronized with the boat NTP server, their screen would allow visualizing the data in real-time, and they are widely available in most scientific laboratories. The software TeraTerm was chosen for a first campaign at sea. It proved to be extremely reliable and covered all the specifications required except for the real-time visualization; an attribute we felt was critical to ensure that the instruments operate well. We then adapted an open-source solution to our needs called Instrumentino (*Koenka et al.*, 2014, 2015). We modified this software to make it suitable for long term deployments by improving its memory management (restricting the data buffer to a fixed-size array). This proof-of-concept was successfully tested on a first cruise (transatlantic crossing on the research vessel Tara in 2016). Subsequently, we sought to design an open-source code, modular (seven unique instruments implemented to date), and robust (logs without user interaction for >30 days) data logging solution with real-time visualization for in situ ocean optical instrumentation: Inlinino. In addition to the software, a data acquisition (DAQ) device for interfacing instruments with analog outputs to computers, was built in an open-source philosophy and is also presented here. The DAQ is a precision analog to serial RS-232 converter, hereafter referred to as PASC.

5.2 Inlinino Software

The core of Inlinino includes a series of simple features essential to its long-time deployment, making it robust, and minimizing its computational footprint. The flow of data between the instruments and the computer is presented in Figure 5.1. Inlinino is entirely configured (e.g. selected instruments) through a configuration file loaded when the software starts. Incoming data from instruments are first timestamped with the computer time. The time of the computer is set to the UTC time zone and synchronized to the NTP server of the research vessel using the operating system service (e.g. Windows Time Service) ensuring accurate timestamping (within a second). Next, the data are stored in a ring-buffer, similar to a first in first out (FIFO) array except that the size of the array is constant. This enables data visualization of the last few minutes (using the fast PyQtGraph library, pyqtgraph.org) and control of the software memory footprint. The data are then written to the computer's storage system (e.g. hard drive) every minute into comma-separated values (CSV) files, limiting the loss of data in case of unexpected computer failure (e.g. power failure). The size of each file generated is constrained to one hour of records, restricting the file size and preventing the creation of unpractical oversized files. The directory in which the data are written is shared on the local network, with read-only permission, enabling automatic daily backup from other computers using synchronization software (e.g. SyncToy on Windows and rsync for Linux and macOS).

At the time of writing, the instruments available include any analog instrument up to 16-bit precision with a voltage ranging from 0 to 6.144 V (using the PASC or a DATAQ DI-1100), such as WET Labs WETStar fluorometers, and the following serial instruments: WET Labs AC-9, WET Labs AC-S, WET Labs ECO-Triplets (e.g. BB3, BBFL2), WET Labs BB9, WET Labs BBRT, WET Labs UBAT, Sea-Bird SBE45 thermosalinograph, Sequoia LISST-100X and Biospherical PAR sensor. Inlinino is modular and designed to interface with any type of serial instrument; therefore, users with good Python knowledge can add new instruments.

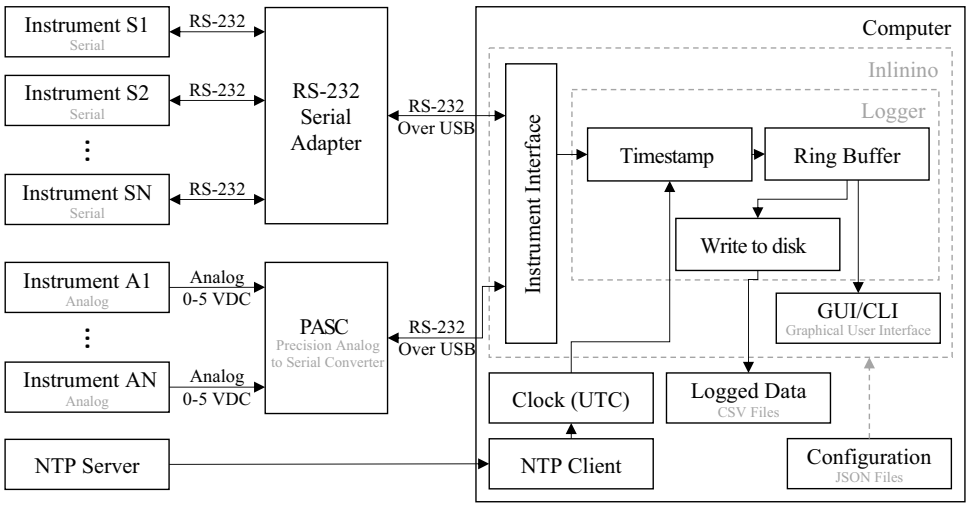

Figure 5.1. Flow chart representation of Inlinino software running on a computer and associated instruments.

5.3 Materials and Instructables

Inlinino hardware and software requirements are described below. Inlinino software was tested on Windows 7 and 10, macOS 10.12, 10.13, and 10.15 operating systems but should run on any environment supporting Python 3 and pyQt 5 (e.g. Windows >7, macOS >10.15, and Linux). The computer running Inlinino requires as many serial ports and analog input channels as the number of instruments with serial and analog communication used. A serial to USB converter is likely needed as most modern computers rarely ship with serial ports. We had good experiences with adapters from the following brands: Aten, Moxa, and StarTech. Regarding the analog input channels, a DAQ device compatible with Inlinino is required only if data from analog instruments need to be logged. To date, the PASC and the DATAQ DI-1100 are the only DAQ devices supported by Inlinino but others could be implemented (e.g. Measurement Computing USB-201, National Instruments USB-6000, Omega USB-1208FS).

Inlinino's installation and first use instructions are briefly described below, for a more detailed documentation we recommend using the instructions available on ReadTheDocs (https://inlinino.readthedocs.io/). Inlinino executable for Windows and macOS can be

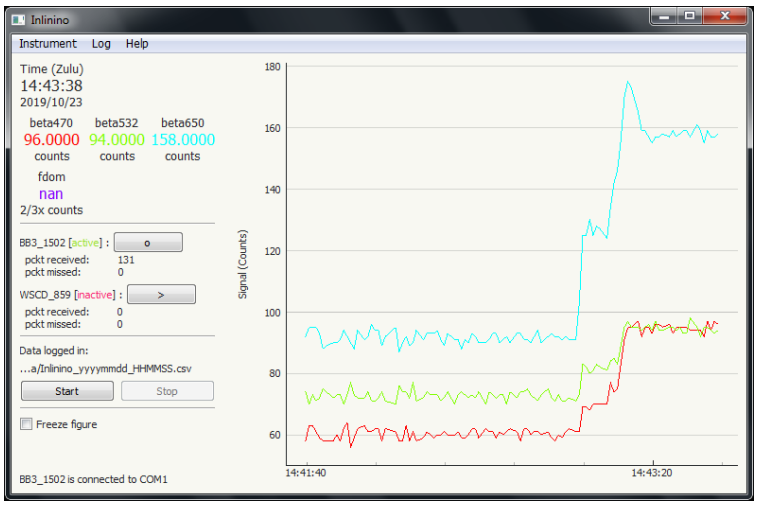

Figure 5.2. Inlinino graphical user interface (GUI).

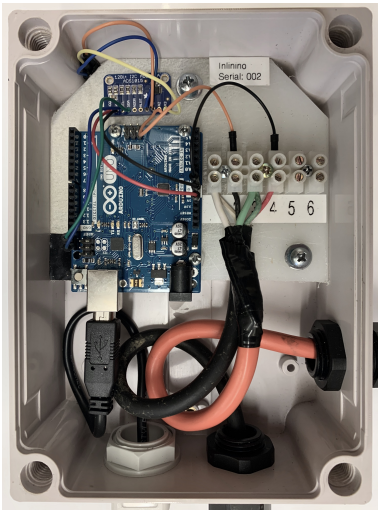

Figure 5.3. Assembly of the precision analog-to-serial converter (PASC) in a waterproof enclosure rated IP66.

downloaded from (ftp://misclab.umeoce.maine.edu/software/Inlinino/). For other operating systems (Linux) Inlinino must be installed from the source code, the procedure and codes are available on GitHub (`https://github.com/OceanOptics/Inlinino`). Once installed instruments parameters (e.g. instruments to log, instruments type) can be setup through the window poping-up at Inlinino start. The application can be started by double-clicking on its icon. The graphical user interface (GUI) of Inlinino (Figure 5.2) is designed to be intuitive allowing experienced users to skip the user manual readings.

Table 5.1. Parts used to build a PASC. *(a)* First version built for $56.90. *(b)* Smaller version built for $28.40.

(a) Components	Price (USD)
Arduino Uno	$22.00
Adafruit ADS1015	$9.95
Waterproof Enclosure IP66	$19.95
Barrier Strips + Electrical Wires	<$5.00

(b) Components	Price (USD)
Adafruit Trinket M0	$8.95
Adafruit ADS1015	$9.95
Aluminum Enclosure IP54	$4.50
Barrier Strips + Electrical Wires	<$5.00

For interfacing analog instruments to the computer running Inlinino, we needed a DAQ device capable of converting the analog signal to a digital signal. In the open-source philosophy of Inlinino, we built a simple precision analog to serial converter (PASC), primarily for educational and low-cost purposes. The PASC (Figure 5.3) is composed of a precision analog to digital converter (ADC, Texas Instrument ADS1015 or ADS1115) and a microcontroller (Arduino Uno) "translating" the digital signal (I2C protocol) from the ADC to serial communication over USB. Those components are packed into a water resistant box (for a complete parts list see Table 5.1). To further reduce the cost of the PASC, the Arduino Uno could be replaced by a cheaper microcontroller that has two serial ports: supporting I2C and UART that could be shifted to RS-232 level or directly converted to a virtual serial port through USB (Table 5.1.b). The wiring of the PASC is explained on Adafruit ADS1015's web page (`https://learn.adafruit.com/adafruit-4-channel-adc-breakouts/assembly-and-wiring`) and the firmware to upload on the microcontroller is available on the GitHub repository of Inlinino (`https://github.com/OceanOptics/Inlinino`).

5.4 Validation and Field Application

To validate the behavior of Inlinino, we compared the logged data to a proven serial terminal emulator TeraTerm (downloaded >7.6 million times as of Jan 23, 2020). We ran

an experiment with a WET Labs BB3 sensor (serial number 1502) connected to a "spy" RS232 cable connected to two serial ports of a computer. A "spy" RS232 cable consists of two receivers for the same transmitter. Therefore, the exact same data could be received at the same time by both TeraTerm and Inlinino. The experiment ran for four hours and 16 minutes. We then analyzed the data to check the following points: (1) dropout rate: does the software miss data?, (2) timestamp accuracy: how accurate is the timestamp?, and (3) data quality: does the data received match the data sent by the sensor?

1. Dropout rate: the exact same number of data lines (n=13596) were received and written to disk on both TeraTerm and Inlinino.

2. Timestamp accuracy: the timestamp of Inlinino was accurate within 1 second with an average delay of 0.54 seconds and 75^{th} percentile at 0.8 seconds. This lag can be explained by the fact that Inlinino writes data from all the instruments it is connected to every N seconds (setup in the configuration of Inlinino). For oceanographic underway measurements, this is acceptable as most measurements are averaged over 30 seconds or more.

3. Data quality: identical values were written to disk for each channel of the BB3 sensor.

The data visualization of Inlinino was qualitatively assessed during the same experiment. The data displayed in TeraTerm was the same as the data plotted in Inlinino. A short delay (<1 second) was noticed on Inlinino's interface. As Inlinino uses the same buffer to write the data to disk and render the visualization, the visualization was assumed to be correct. In addition, Inlinino was deployed on several research vessels (Table 5.2) during more than 650 days at sea. For most of the campaigns, the data were manually quality checked (see Boss et al. 2019) and published in NASA SeaBASS database.

The computing performance of Inlinino was assessed by running it with one instrument (WET Labs BB3 serial number 1502) on a regular laptop (500 $, Lenovo E530, CPU i3-3120 @2.5 GHz, 4Gb RAM, Windows 7). The CPU usage was monitored for two hours

Table 5.2. List of campaigns on which the Inlinino software was deployed. The precision analog to serial converter was deployed on each campaign including a WSCD sensor. (*)Note that during the Tara Pacific Expedition, the instruments were installed at different times and had to be shipped for maintenance, resulting in different numbers of days logged between instruments.

(b) Campaign	Date	Instruments Logged	Days Logging
NAAMES II	May 2016	BB3 & WSCD	26
Schiller Coastal Studies Center	Summer 2016	TSG, BB3, BBFL2, & 3X1M	90
PEACETIME	May 2017	FLBBCD, BB3, & WSCD	33
NAAMES III	Sep 2017	BB3 & WSCD	26
SPURS-2	Oct-Nov 2017	BB3	30
NAAMES IV	Mar-Jun 2018	BB3 & WSCD	26
EXPORTS I: *R/V Roger Revelle*	Aug-Sep 2018	BB3 & WSCD	34
EXPORTS I: *R/V Sally Ride*	Aug-Sep 2018	TSG, BB3, BBFL2, & 3X1M	36
Tara Pacific	May 2016-Oct 2018	PAR, BB3, & WSCD	495, 261, & 373*
Tara Breizh Bloom	May 2019	BB3 & WSCD	7

and five minutes with the Windows Performance Monitor utility. Inlinino was the only application running during the test. The processor's clock frequency was automatically down-throttled to 1.3 GHz 66 % of the time with a median CPU usage of 1.2 % and 75^{th} percentile of 2.0 %, validating its low computing power characteristics. Inlinino was also used on several laptops along multiple month campaigns (Table 5.2): CPU usage was minimal (qualitatively assessed through Windows Task Manager) and no problems of data logging were encountered.

The accuracy of the PASC was first assessed in the laboratory when it was built and during its first campaigns at sea with a Fluke 85 III voltmeter. While it performed as expected (close to the Texas Instrument ADS1015 specifications), no data were kept for a scientific evaluation of its performance. The accuracy and precision of the PASC serial number 001 and 002 were evaluated with a Fluke 85 III voltmeter, after 145 and 380 days of operation at sea (Figure available on Inlinino's documentation, https://inlinino.readthedocs.io/en/latest/pasc.html). The WET Labs WETStar WSCD serial number 859 was connected to the PASC for the test and the full range of values

output by this sensor (0 to 5 V) was used. Difference between the two PASC and the reference voltmeter was never greater than 0.05 V with a root mean squared error of 0.0053 V. While no significant bias was observed, the precision was twice higher than the manufacturer specification (0.003 V). For other characteristics of the electronic chip used in the PASC such as drift over temperature (0.015 mV/°C), we refer the user to the manufacturer datasheet.

5.5 Summary and Future Work

The current version of Inlinino is operational, robust, easy to use on a routine basis while at sea. It allows for the smooth collection of days of data at a time and fits the needs for logging measurements with underway systems. As Inlinino is shared and open-sourced it can be adjusted as needed.

From our experience using Inlinino, we identified a few features to improve

- Several times the instruments were running, and data were displayed in real-time, but the logging was disabled, so the data were not recorded (about once per expedition). We concluded that this problem could be reduced by including a large visual and/or audio signal in the interface, warning the user when data are not being logged.

- The configuration of the instruments at the beginning of each campaign was not intuitive and could be difficult to make for users with limited coding skills. With this in mind, we thought that it would be good to add a user interface to configure instruments at the beginning of each campaign.

Future development of Inlinino could also include

- Enable state-of-the-art processing, applying optical corrections, and computing products such as phytoplankton carbon or chlorophyll (*Boss et al.*, 2019) in near real-time. The processing algorithms were developed and tested on most of the campaigns mentioned above; however, the graphical user interface currently provided

Figure 5.4. Biological Oceanography Class (SMS 501) from the University of Maine observing chlorophyll a fluorescence as they progress through the Damariscotta (Maine) Estuary (Maine, USA) on R/V Ira C. Measurements from the WET Labs WETStar WS3S fluorometer and the Sea-Bird SBE 16plus CTD sensor were logged by Inlinino and ThingsBoard running on a single board computer (Raspberry Pi 3) and streamed to a tablet in real time. (c) Guillaume Bourdin

is limited, limiting the utility of the software (InlineAnalysis, https://github.com/OceanOptics/InLineAnalysis). The real-time processing inside Inlinino would enable troubleshooting earlier in the acquisition of data thus helping scientists to make data-informed decisions in near-real-time in the field.

- Streaming in real-time data recorded by Inlinino to any web browser, including phones and tablets. This could provide monitoring of the system from anywhere on the research vessel while at sea. Recently, for use in a field class, we developed a first proof of concept (Figure 5.4) which takes advantage of the recent development of the Internet of Things (IoT) and ThingsBoard platform (https://thingsboard.io). The source code of this module is available on the dedicated branch of Inlinino on GitHub (https://github.com/OceanOptics/Inlinino/tree/tb-app).

CHAPTER 6

CONCLUSION

The present thesis showcases efforts toward understanding the role of phytoplankton in inherent optical properties. The four major thrusts to this research were: (1) build a phytoplankton size distribution (PSD) by merging measurements from two cytometers; (2) evaluate current phytoplankton size proxies; (3) extend size proxies to larger organisms responsible for layers of spike in backscattering profiles; and (4) estimate the contribution of phytoplankton to particulate attenuation and backscattering. In the process, a robust software data logger was developed to record data from optical instrumentation that continuously collects measurements of surface ocean optical properties.

The phytoplankton size distribution covered a broad set of biological and physical conditions spanning four seasons across the western North Atlantic from the subtropic to the subtropical gyre. To my knowledge, these results present the first phytoplankton size distribution extended over a broad size range in the western North Atlantic and supplement a scarce dataset of phytoplankton size distribution. The trends in the size distribution aligned well with current understanding of the phytoplankton annual cycle. PSDs are dominated by picoplankton in the subtropical gyre throughout the year while having an increased contribution by small nano-phytoplankton (< 8 µm) in spring. In northern subregions, there was a shift in size throughout the year. In the spring, larger phytoplankton dominated the community transitioning to small nanophytoplankton throughout the summer. In the early winter, phytoplankton size groups were evenly distributed and associated with low concentrations. Following the recommendations of *Jonasz and Fournier* (2007), I quantified the uncertainties associated with the PSD build. I included all uncertainty sources that could be identified, from the precision of the cytometers to counting statistics, as well as orientation of the cells *in-situ* vs in-cytometers. One of the major challenges in the process was calibrating the cytometers. In fact, there

were inconsistencies when measuring similar samples with two different cytometers (*Lombard et al.*, 2019).

A variety of algorithms which provide information on phytoplankton size were tested with the phytoplankton size distribution derived here. Surprisingly, I found chlorophyll *a* concentration to correlate better with observed size parameters of phytoplankton than did most algorithms dedicated to its retrieval using other optically based proxies. Nonetheless, the absorption band effect and spectral slope of attenuation were the best predictors of the phytoplankton size distribution. This suggests that algorithm development efforts associated with ocean color, should focus on obtaining the backscattering spectra with better accuracy than currently done (*Bisson et al.*, 2019).

The phytoplankton and particulate size distribution coincident with the optical observations revealed that phytoplankton dominate the particulate attenuation and backscattering in the open ocean across a wide range of conditions ($\sim$ 75 %). This is significantly more than previously reported. Hence, it reinforces the idea that particulate attenuation and backscattering are robust predictors of phytoplankton biomass and other parameters such as their size. The present study provides a unique dataset to test novel models taking into account the complexity of marine particles.

Using a novel spike layer detection method (developed in Chapter 4), I could distinguish the origin of the spikes in backscattering and FDOM profiles collected with profiling floats. This method can be used to refine estimation of carbon export due to marine snow and detect mesopelagic organisms for the first time with autonomous platforms already deployed throughout the ocean on biogeochemical-Argo floats. Furthermore, this type of analysis provides a unique opportunity to study the distribution of these organisms playing an important role in the biological pump and extend observations made by more "classical" methodology (e.g. nets and acoustic aboard research vessels).

During my thesis, the software data logger I developed, Inlinino, recorded more than 650 days of measurements from a multitude of optical sensors (e.g. WET Labs ECO-BB3,

WSCD), on six research vessels. It drastically improved the ability to record and quality check data acquired with commercial optical instruments on month long campaigns in real time. This task was rather complicated with exiting commercial solutions not designed for such extended deployments. The data collected with Inlinino was used in the analysis of this thesis and was shared on NASA SeaBASS with the oceanographic community. Inlinino now also supports high throughput from hyperspectral sensors such as the WET Labs ACS and Satlantic HyperOCR sensors. It has been shared with the community (`https://github.com/OceanOptics/Inlinino`) and is ready to be used on board new expeditions, the next one being with R/V Tara beginning December 2020.

REFERENCES

Aas, E. (1996), Refractive index of phytoplankton derived from its metabolite composition, *Journal of Plankton Research*, *18*(12), 2223–2249, doi:10.1093/plankt/18.12.2223.

Abràmoff, M. D., P. J. Magalhães, and S. J. Ram (2004), Image processing with imageJ, *Biophotonics International*, *11*(7), 36–41.

Baetge, N., J. R. Graff, M. J. Behrenfeld, and C. A. Carlson (2020), Net Community Production, Dissolved Organic Carbon Accumulation, and Vertical Export in the Western North Atlantic, *Frontiers in Marine Science*, *7*, 227, doi:10.3389/fmars.2020.00227.

Bax, N. J., P. Miloslavich, F. E. Muller-Karger, V. Allain, W. Appeltans, S. D. Batten, L. Benedetti-Cecchi, P. L. Buttigieg, S. Chiba, D. P. Costa, J. E. Duffy, D. C. Dunn, C. R. Johnson, R. M. Kudela, D. Obura, L.-M. Rebelo, Y.-J. Shin, S. E. Simmons, and P. L. Tyack (2019), A Response to Scientific and Societal Needs for Marine Biological Observations, *Frontiers in Marine Science*, *6*, 395, doi:10.3389/fmars.2019.00395.

Behrenfeld, M. J., and E. Boss (2006), Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass, *Journal of Marine Research*, *64*(3), 431–451, doi:10.1357/002224006778189563.

Behrenfeld, M. J., and E. S. Boss (2018), Student's tutorial on bloom hypotheses in the context of phytoplankton annual cycles, *Global Change Biology*, *24*(1), 55–77, doi:10.1111/gcb.13858.

Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea (2005), Carbon-based ocean productivity and phytoplankton physiology from space: Phytoplankton Growth Rates and Ocean Productivity, *Global Biogeochemical Cycles*, *19*(1), doi:10.1029/2004GB002299.

Behrenfeld, M. J., R. H. Moore, C. A. Hostetler, J. Graff, P. Gaube, L. M. Russell, G. Chen, S. C. Doney, S. Giovannoni, H. Liu, C. Proctor, L. M. Bolaños, N. Baetge, C. Davie-Martin, T. K. Westberry, T. S. Bates, T. G. Bell, K. D. Bidle, E. S. Boss, S. D. Brooks, B. Cairns, C. Carlson, K. Halsey, E. L. Harvey, C. Hu, L. Karp-Boss, M. Kleb, S. Menden-Deuer, F. Morison, P. K. Quinn, A. J. Scarino, B. Anderson, J. Chowdhary, E. Crosbie, R. Ferrare, J. W. Hair, Y. Hu, S. Janz, J. Redemann, E. Saltzman, M. Shook, D. A. Siegel, A. Wisthaler, M. Y. Martin, and L. Ziemba (2019), The North Atlantic Aerosol and Marine Ecosystem Study (NAAMES): Science Motive and Mission Overview, *Frontiers in Marine Science*, *6*, doi:10.3389/fmars.2019.00122.

Benoit-Bird, K. J., T. Patrick Welch, C. M. Waluk, J. A. Barth, I. Wangen, P. McGill, C. Okuda, G. A. Hollinger, M. Sato, and S. McCammon (2018), Equipping an underwater glider with a new echosounder to explore ocean ecosystems, *Limnology and Oceanography: Methods*, *16*(11), 734–749, doi:10.1002/lom3.10278.

Bergquist, A. M., S. R. Carpenter, and J. C. Latino (1985), Shifts in phytoplankton size structure and community composition during grazing by contrasting zooplankton assemblages1: Phytoplankton size structure, *Limnology and Oceanography*, *30*(5), 1037–1045, doi:10.4319/lo.1985.30.5.1037.

Bertilsson, S., O. Berglund, D. M. Karl, and S. W. Chisholm (2003), Elemental composition of marine Prochlorococcus and Synechococcus : Implications for the ecological stoichiometry of the sea, *Limnology and Oceanography*, *48*(5), 1721–1731, doi:10.4319/lo.2003.48.5.1721.

Bianchi, D., and K. A. S. Mislan (2016), Global patterns of diel vertical migration times and velocities from acoustic data, *Limnology and Oceanography*, *61*(1), 353–364, doi:10.1002/lno.10219.

Bishop, J. K., and T. Wood (2008), Particulate matter chemistry and dynamics in the twilight zone at VERTIGO ALOHA and K2 sites, *Deep Sea Research Part I: Oceanographic Research Papers*, *55*(12), 1684–1706, doi:10.1016/j.dsr.2008.07.012.

Bishop, J. K., S. E. Calvert, and M. Y. Soon (1999), Spatial and temporal variability of POC in the northeast Subarctic Pacific, *Deep Sea Research Part II: Topical Studies in Oceanography*, *46*(11-12), 2699–2733, doi:10.1016/S0967-0645(99)00081-8.

Bisson, K. M., E. Boss, T. K. Westberry, and M. J. Behrenfeld (2019), Evaluating satellite estimates of particulate backscatter in the global open ocean using autonomous profiling floats, *Optics Express*, *27*(21), 30,191, doi:10.1364/OE.27.030191.

Bittig, H., A. Wong, J. Plant, and CORIOLIS-ADMT (2018), BGC-Argo synthetic profile file processing and format on Coriolis GDAC, *Tech. rep.*, doi:http://dx.doi.org/10.13155/55637.

Bohren, C., and D. Huffman (1983), *Absorption and Scattering of Light by Small Particles*, A Wiley-Interscience publication, Wiley.

Bolaños, L. M., L. Karp-Boss, C. J. Choi, A. Z. Worden, J. R. Graff, N. Haëntjens, A. P. Chase, A. Della Penna, P. Gaube, F. Morison, S. Menden-Deuer, T. K. Westberry, R. T. O'Malley, E. Boss, M. J. Behrenfeld, and S. J. Giovannoni (2020), Small phytoplankton dominate western North Atlantic biomass, *The ISME Journal*, *14*(7), 1663–1674, doi:10.1038/s41396-020-0636-0.

Boss, E., W. S. Pegau, W. D. Gardner, J. R. V. Zaneveld, A. H. Barnard, M. S. Twardowski, G. C. Chang, and T. D. Dickey (2001a), Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental shelf, *Journal of Geophysical Research: Oceans*, *106*(C5), 9509–9516, doi:10.1029/2000JC900077.

Boss, E., M. S. Twardowski, and S. Herring (2001b), Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution, *Applied Optics*, *40*(27), 4885, doi:10.1364/AO.40.004885.

Boss, E., W. H. Slade, M. Behrenfeld, and G. Dall'Olmo (2009), Acceptance angle effects on the beam attenuation in the ocean, *Optics Express*, *17*(3), 1535, doi:10.1364/OE.17.001535.

Boss, E., L. Guidi, M. J. Richardson, L. Stemmann, W. Gardner, J. K. B. Bishop, R. F. Anderson, and R. M. Sherrell (2015), Optical techniques for remote and in-situ characterization of particles pertinent to GEOTRACES, *Progress in Oceanography*, *133*, 43–54, doi:10.1016/j.pocean.2014.09.007.

Boss, E., A. Waite, F. Muller-Karger, H. Yamazaki, R. Wanninkhof, J. Uitz, S. Thomalla, H. Sosik, B. Sloyan, A. Richardson, P. Miloslavich, J. Karstensen, G. Grégori, K. Fennel, H. Claustre, M. Cornejo, I. Berman-Frank, S. Batten, and S. Acinas (2018a), Beyond Chlorophyll Fluorescence: The Time is Right to Expand Biological Measurements in Ocean Observing Programs, *Limnology and Oceanography Bulletin*, *27*(3), 89–90, doi:10.1002/lob.10243.

Boss, E., N. Haëntjens, T. K. Westberry, L. Karp-Boss, and W. H. Slade (2018b), Validation of the particle size distribution obtained with the laser in-situ scattering and transmission (LISST) meter in flow-through mode, *Optics Express*, *26*(9), 11,125, doi:10.1364/OE.26.011125.

Boss, E., N. Haëntjens, S. Ackleson, B. Balch, A. Chase, G. D. Olmo, S. Freeman, Y. Liu, J. Loftin, W. Neary, N. Nelson, M. Novak, W. Slade, C. Proctor, P. Tortell, and T. Westberry (2019), Inherent Optical Property Measurements and Protocols: Best practices for the collection and processing of ship-based underway flow-through optical data, *Tech. rep.*, IOCCG.

Boss, E. S., R. Collier, G. Larson, K. Fennel, and W. S. Pegau (2007a), Measurements of spectral optical properties and their relation to biogeochemical variables and processes in Crater Lake, Crater Lake National Park, OR, *Hydrobiologia*, *574*(1), 149–159, doi:10.1007/s10750-006-2609-3.

Boss, E. S., R. Collier, G. Larson, K. Fennel, and W. S. Pegau (2007b), Measurements of spectral optical properties and their relation to biogeochemical variables and processes in Crater Lake, Crater Lake National Park, OR, *Hydrobiologia*, *574*(1), 149–159, doi:10.1007/s10750-006-2609-3.

Box, G. E. P., and D. R. Cox (1964), An Analysis of Transformations, *Journal of the Royal Statistical Society*, *26*(2), 211–252.

Boyd, P. W., H. Claustre, M. Levy, D. A. Siegel, and T. Weber (2019), Multi-faceted particle pumps drive carbon sequestration in the ocean, *Nature*, *568*(7752), 327–335, doi:10.1038/s41586-019-1098-2.

Bricaud, A. (2004), Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations, *Journal of Geophysical Research*, *109*(C11), C11,010, doi:10.1029/2004JC002419.

Bricaud, A., A.-L. Bédhomme, and A. Morel (1988), Optical properties of diverse phytoplanktonic species: experimental results and theoretical interpretation, *Journal of Plankton Research*, *10*(5), 851–873, doi:10.1093/plankt/10.5.851.

Briggs, N., M. J. Perry, I. Cetinić, C. Lee, E. D'Asaro, A. M. Gray, and E. Rehm (2011), High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom, *Deep Sea Research Part I: Oceanographic Research Papers*, *58*(10), 1031–1039, doi:10.1016/j.dsr.2011.07.007.

Brown, C. M., J. E. Lawrence, and D. A. Campbell (2006), Are phytoplankton population density maxima predictable through analysis of host and viral genomic DNA content?, *Journal of the Marine Biological Association of the United Kingdom*, *86*(3), 491–498, doi:10.1017/S0025315406013397.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West (2004), Toward a metabolic theory of ecology., *Ecology*, *85*(7), 1771–1789, doi:10.1890/03-9000.

Brown, O. B., and H. R. Gordon (1974), Size–Refractive Index Distribution of Clear Coastal Water Particulates from Light Scattering, *Applied Optics*, *13*(12), 2874, doi:10.1364/AO.13.002874.

Burt, W. J., and P. D. Tortell (2018), Observations of Zooplankton Diel Vertical Migration From High-Resolution Surface Ocean Optical Measurements, *Geophysical Research Letters*, *45*(24), 396–13, doi:10.1029/2018GL079992.

Carroll, M., D. Chigounis, S. Gilbert, K. Gundersen, K. Hayashi, C. Janzen, T. Johengen, T. Koles, F. Laurier, T. McKissack, L. Meadows, C. Metcalfe, C. Robertson, D. Schar, J. Seiter, G. J. Smith, M. Tamburri, and D. Wells (2005), Performance Verification Statement for the WET Labs ECO FLNTUSB fluorometer, *Tech. Rep. ACT VS07-06*, doi:http://dx.doi.org/10.25607/OBP-372.

Cauchy, A. L. (1832), *Mémoire sur la rectification des courbes et la quadrature des surfaces courbes*.

Cetinić, I., G. Toro-Farmer, M. Ragan, C. Oberg, and B. H. Jones (2009), Calibration procedure for Slocum glider deployed optical instruments, *Optics Express*, *17*(18), 15,420, doi:10.1364/OE.17.015420.

Cetinić, I., M. J. Perry, N. T. Briggs, E. Kallin, E. A. D'Asaro, and C. M. Lee (2012), Particulate organic carbon and inherent optical properties during 2008 North Atlantic bloom experiment, *Journal of Geophysical Research: Oceans*, *117*(6), doi:10.1029/2011JC007771.

Cetinić, I., N. Poulton, and W. H. Slade (2016), Characterizing the phytoplankton soup: pump and plumbing effects on the particle assemblage in underway optical seawater systems, *Optics Express*, *24*(18), 20,703, doi:10.1364/OE.24.020703.

Chai, F., K. S. Johnson, H. Claustre, X. Xing, Y. Wang, E. Boss, S. Riser, K. Fennel, O. Schofield, and A. Sutton (2020), Monitoring ocean biogeochemistry with autonomous platforms, *Nature Reviews Earth & Environment*, *1*(6), 315–326, doi:10.1038/s43017-020-0053-y.

Chase, A., E. Boss, R. Zaneveld, A. Bricaud, H. Claustre, J. Ras, G. Dall'Olmo, and T. K. Westberry (2013), Decomposition of in situ particulate absorption spectra, *Methods in Oceanography*, *7*, 110–124, doi:10.1016/j.mio.2014.02.002.

Chase, A. P., S. J. Kramer, N. Haëntjens, E. S. Boss, L. Karp-Boss, M. Edmondson, and J. R. Graff (2020), Evaluation of diagnostic pigments to estimate phytoplankton size classes, *Limnology and Oceanography: Methods*, p. lom3.10385, doi:10.1002/lom3.10385.

Chisholm, S. W. (1992), Phytoplankton Size, in *Primary Productivity and Biogeochemical Cycles in the Sea*, edited by P. G. Falkowski, A. D. Woodhead, and K. Vivirito, pp. 213–237, Springer US, Boston, MA, doi:10.1007/978-1-4899-0762-2_12.

Ciotti, \. M., M. R. Lewis, and J. J. Cullen (2002), Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient, *Limnology and Oceanography*, *47*(2), 404–417, doi:10.4319/lo.2002.47.2.0404.

Ciotti, A. M., and A. Bricaud (2006), Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil: Algal size and CDM from SeaWiFS data, *Limnology and Oceanography: Methods*, *4*(7), 237–253, doi:10.4319/lom.2006.4.237.

Claustre, H., K. S. Johnson, and Y. Takeshita (2020), Observing the Global Ocean with Biogeochemical-Argo, *Annual Review of Marine Science*, *12*(1), 010,419–010,956, doi:10.1146/annurev-marine-010419-010956.

Clavano, W., E. Boss, and L. Karp-Boss (2007), Inherent Optical Properties of Non-Spherical Marine-Like Particles – From Theory To Observation, in *Oceanography and Marine Biology*, vol. 20074975, edited by R. Gibson, R. Atkinson, and J. Gordon, pp. 1–38, CRC Press, doi:10.1201/9781420050943.ch1, series Title: Oceanography and Marine Biology - An Annual Review.

Colebrook, J. (1982), Continuous plankton records: seasonal variations in the distribution and abundance of plankton in the North Atlantic Ocean and the North Sea, *Journal of Plankton Research*, *4*(3), 435–462, doi:10.1093/plankt/4.3.435.

Dall'Olmo, G., and K. A. Mork (2014), Carbon export by small particles in the Norwegian Sea, *Geophysical Research Letters*, *41*(8), 2921–2927, doi:10.1002/2014GL059244.

Dall'Olmo, G., T. K. Westberry, M. J. Behrenfeld, E. Boss, and W. H. Slade (2009), Significant contribution of large particles to optical backscattering in the open ocean, *Biogeosciences Discussions*, *6*(1), 291–340, doi:10.5194/bgd-6-291-2009.

Davis, R. F., C. C. Moore, J. R. V. Zaneveld, and J. M. Napp (1997), Reducing the effects of fouling on chlorophyll estimates derived from long-term deployments of optical instruments, *Journal of Geophysical Research: Oceans*, *102*(C3), 5851–5855, doi:10.1029/96JC02430.

De Robertis, A., and N. O. Handegard (2013), Fish avoidance of research vessels and the efficacy of noise-reduced vessels: a review, *ICES Journal of Marine Science*, *70*(1), 34–45, doi:10.1093/icesjms/fss155.

De Robertis, A., and I. Higginbottom (2007), A post-processing technique to estimate the signal-to-noise ratio and remove echosounder background noise, *ICES Journal of Marine Science*, *64*(6), 1282–1291, doi:10.1093/icesjms/fsm112.

De Robertis, A., N. Lawrence-Slavas, R. Jenkins, I. Wangen, C. W. Mordy, C. Meinig, M. Levine, D. Peacock, and H. Tabisola (2019), Long-term measurements of fish backscatter from Saildrone unmanned surface vehicles and comparison with observations from a noise-reduced research vessel, *ICES Journal of Marine Science*, doi:10.1093/icesjms/fsz124.

Della Penna, A., and P. Gaube (2019), Overview of (Sub)mesoscale Ocean Dynamics for the NAAMES Field Program, *Frontiers in Marine Science*, *6*, 384, doi:10.3389/fmars.2019.00384.

DeLong, J. P., and D. A. Vasseur (2012), Size-density scaling in protists and the links between consumer-resource interaction parameters, *Journal of Animal Ecology*, *81*(6), 1193–1201, doi:10.1111/j.1365-2656.2012.02013.x.

Devred, E., S. Sathyendranath, V. Stuart, and T. Platt (2011), A three component classification of phytoplankton absorption spectra: Application to ocean-color data, *Remote Sensing of Environment*, *115*(9), 2255–2266, doi:10.1016/j.rse.2011.04.025.

Dickey, T. D. (1991), The emergence of concurrent high-resolution physical and bio-optical measurements in the upper ocean and their applications, *Reviews of Geophysics*, *29*(3), 383, doi:10.1029/91RG00578.

Dierssen, H. M., R. M. Kudela, J. P. Ryan, and R. C. Zimmerman (2006a), Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments, *Limnology and Oceanography*, *51*(6), 2646–2659, doi:10.4319/lo.2006.51.6.2646.

Dierssen, H. M., R. M. Kudela, J. P. Ryan, and R. C. Zimmerman (2006b), Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments, *Limnology and Oceanography*, *51*(6), 2646–2659, doi:10.4319/lo.2006.51.6.2646.

Durand, M. D., and R. J. Olson (1996), Contributions of phytoplankton light scattering and cell concentration changes to diel variations in beam attenuation in the equatorial Pacific from flow cytometric measurements of pico-, ultra- and nanoplankton, *Deep Sea Research Part II: Topical Studies in Oceanography*, *43*(4-6), 891–906, doi:10.1016/0967-0645(96)00020-3.

DuRand, M. D., R. J. Olson, and S. W. Chisholm (2001), Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea, *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(8-9), 1983–2003, doi:10.1016/S0967-0645(00)00166-1.

Estapa, M. L., M. L. Feen, and E. Breves (2019), Direct Observations of Biological Carbon Export From Profiling Floats in the Subtropical North Atlantic, *Global Biogeochemical Cycles*, *33*(3), 282–300, doi:10.1029/2018GB006098.

Falkenhaug, T., K. Tande, and T. Semenova (1997), Diel, seasonal and ontogenetic variations in the vertical distributions of four marine copepods, *Marine Ecology Progress Series*, *149*(1-3), 105–119, doi:10.3354/meps149105.

Font-Muñoz, J. S., R. Jeanneret, J. Arrieta, S. Anglès, A. Jordi, I. Tuval, and G. Basterretxea (2019), Collective sinking promotes selective cell pairing in planktonic pennate diatoms, *Proceedings of the National Academy of Sciences*, *116*(32), 15,997–16,002, doi:10.1073/pnas.1904837116.

Francois, R. E., and G. R. Garrison (1982), Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption, *The Journal of the Acoustical Society of America*, *72*(6), 1879–1890, doi:10.1121/1.388673.

Freedman, D., and P. Diaconis (1981), On the histogram as a density estimator:L 2 theory, *Zeitschrift for Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *57*(4), 453–476, doi:10.1007/BF01025868.

Frouin, R., D. W. Lingner, C. Gautier, K. S. Baker, and R. C. Smith (1989), A simple analytical formula to compute clear sky total and photosynthetically available solar irradiance at the ocean surface, *Journal of Geophysical Research*, *94*(C7), 9731, doi:10.1029/JC094iC07p09731.

Fujiwara, A., T. Hirawake, K. Suzuki, and S.-I. Saitoh (2011), Remote sensing of size structure of phytoplankton communities using optical properties of the Chukchi and Bering Sea shelf region, *Biogeosciences*, *8*(12), 3567–3580, doi:10.5194/bg-8-3567-2011.

Gibbs, R. J. (1981), Floc breakage by pumps, *Journal of Sedimentary Research*, *51*(2), 670–672, doi:10.1306/212F7D56-2B24-11D7-8648000102C1865D.

Gibbs, R. J. (1982a), Floc breakage during HIAC light-blocking analysis, *Environmental Science & Technology*, *16*(5), 298–299, doi:10.1021/es00099a012.

Gibbs, R. J. (1982b), Floc stability during Coulter-counter size analysis, *Journal of Sedimentary Research*, *52*(2), 657–660, doi:10.1306/212F7FE5-2B24-11D7-8648000102C1865D.

Gibbs, R. J., and L. N. Konwar (1983), Notes. Sampling of mineral flocs using Niskin bottles, *Environmental Science & Technology*, *17*(6), 374–375, doi:10.1021/es00112a014.

Giering, S. L. C., B. Hosking, N. Briggs, and M. H. Iversen (2020), The Interpretation of Particle Size, Shape, and Carbon Flux of Marine Particle Images Is Strongly Affected by the Choice of Particle Detection Algorithm, *Frontiers in Marine Science*, *7*, 564, doi:10.3389/fmars.2020.00564.

Gin, K. Y., S. W. Chisholm, and R. J. Olson (1999), Seasonal and depth variation in microbial size spectra at the Bermuda Atlantic time series station, *Deep Sea Research Part I: Oceanographic Research Papers*, *46*(7), 1221–1245, doi:10.1016/S0967-0637(99)00004-7.

Goulet, P., C. Guinet, R. Swift, P. T. Madsen, and M. Johnson (2019), A miniature biomimetic sonar and movement tag to study the biotic environment and predator-prey interactions in aquatic animals, *Deep Sea Research Part I: Oceanographic Research Papers*, *148*(May), 1–11, doi:10.1016/j.dsr.2019.04.007.

Graff, J. R., T. K. Westberry, A. J. Milligan, M. B. Brown, G. Dall'Olmo, V. v. Dongen-Vogels, K. M. Reifel, and M. J. Behrenfeld (2015), Analytical phytoplankton carbon measurements spanning diverse ecosystems, *Deep Sea Research Part I: Oceanographic Research Papers*, *102*, 16–25, doi:10.1016/j.dsr.2015.04.006.

Green, R. E., H. M. Sosik, R. J. Olson, and M. D. DuRand (2003), Flow cytometric determination of size and complex refractive index for marine particles: comparison with independent and bulk estimates, *Applied Optics*, *42*(3), 526, doi:10.1364/AO.42.000526.

Harmel, T., M. Hieronymi, W. Slade, R. Röttgers, F. Roullier, and M. Chami (2016), Laboratory experiments for inter-comparison of three volume scattering meters to measure angular scattering properties of hydrosols, *Optics Express*, *24*(2), A234, doi:10.1364/OE.24.00A234.

Hays, G. C. (2003), A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations, *Hydrobiologia*, *503*(1-3), 163–170, doi:10.1023/B:HYDR.0000008476.23617.b0.

Heath, M. (2000), Winter distribution of Calanus finmarchicus in the Northeast Atlantic, *ICES Journal of Marine Science*, *57*(6), 1628–1635, doi:10.1006/jmsc.2000.0978.

Heldal, M., D. J. Scanlan, S. Norland, F. Thingstad, and N. H. Mann (2003), Elemental composition of single cells of various strains of marine Prochlorococcus and Synechococcus using X-ray microanalysis, *Limnology and Oceanography*, *48*(5), 1732–1743, doi:10.4319/lo.2003.48.5.1732.

Hilligsøe, K. M., K. Richardson, J. Bendtsen, L.-L. Sørensen, T. G. Nielsen, and M. M. Lyngsgaard (2011), Linking phytoplankton community size composition with temperature, plankton food web structure and sea–air CO2 flux, *Deep Sea Research Part I: Oceanographic Research Papers*, *58*(8), 826–838, doi:10.1016/j.dsr.2011.06.004.

Hirata, T., J. Aiken, N. Hardman-Mountford, T. Smyth, and R. Barlow (2008), An absorption model to determine phytoplankton size classes from satellite ocean colour, *Remote Sensing of Environment*, *112*(6), 3153–3159, doi:10.1016/j.rse.2008.03.011.

Hooker, S. B., L. Van Heukelem, C. Thomas, H. Claustre, J. Ras, L. Schluter, L. Clementson, D. Van Der Linde, E. Eker-Develi, J.-F. Berthon, R. Barlow, H. Sessions, H. Ismail, and J. Perl (2009), The Third SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-3), *NASA Tech. Memo,Goddard Space Flight Center*, *2009-21584*.

Houskeeper, H. F., D. Draper, R. M. Kudela, and E. Boss (2020), Chlorophyll absorption and phytoplankton size information inferred from hyperspectral particulate beam attenuation, *Applied Optics*, *59*(22), 6765, doi:10.1364/AO.396832.

Huete-Ortega, M., A. Calvo-Díaz, R. Graña, B. Mouriño-Carballido, and E. Marañón (2011), Effect of environmental forcing on the biomass, production and growth rate of size-fractionated phytoplankton in the central Atlantic Ocean, *Journal of Marine Systems*, *88*(2), 203–213, doi:10.1016/j.jmarsys.2011.04.007.

Ignatyev, S. M. (1997), Pelagic Fishes and Their Macroplankton Prey: Swimming Speeds, in *Forage Fishes in Marine Ecosystems, American Fisheries Society*, 97-01, pp. 31–39, University of Alaska Fairbanks.

Jennings, B. R., and K. Parslow (1988), Particle size measurement: the equivalent spherical diameter, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, *419*(1856), 137–149, doi:10.1098/rspa.1988.0100.

Jerlov, N. G. (1976), *Marine optics*, Elsevier.

Johnson, P. W., and J. M. Sieburth (1979), Chroococcoid cyanobacteria in the sea: A ubiquitous and diverse phototrophic biomass1, *Limnology and Oceanography*, *24*(5), 928–935, doi:10.4319/lo.1979.24.5.0928.

Jonasz, M. (1983), Particle-size distributions in the Baltic, *Tellus B*, *35B*(5), 346–358, doi:10.1111/j.1600-0889.1983.tb00039.x.

Jonasz, M. (1987), Nonsphericity of suspended marine particles and its influence on light scattering, *Limnology and Oceanography*, *32*(5), 1059–1065, doi:10.4319/lo.1987.32.5.1059.

Jonasz, M., and G. R. Fournier (2007), *Light scattering by particles in water: theoretical and experimental foundations*, 1. ed ed., Elsevier, Acad. Press, Amsterdam, oCLC: 255582374.

Junge, C. E. (1963), Air chemistry and radioactivity, *1963*, p. 382.

Karp-Boss, L., and P. A. Jumars (1998), Motion of diatom chains in steady shear flow, *Limnology and Oceanography*, *43*(8), 1767–1773, doi:10.4319/lo.1998.43.8.1767.

Karp-Boss, L., E. Boss, and P. A. Jumars (1996), Nutrient fluxes to planktonic osmotrophs in the presence of fluid motion, *Oceanography and Marine Biology*, *34*, 71–108.

Karp-Boss, L., E. Boss, and P. A. Jumars (2000), Motion of dinoflagellates in a simple shear flow, *Limnology and Oceanography*, *45*(7), 1594–1602, doi:10.4319/lo.2000.45.7.1594.

Karp-Boss, L., L. Azevedo, and E. Boss (2007), LISST-100 measurements of phytoplankton size distribution: evaluation of the effects of cell shape, *Limnology and Oceanography: Methods*, *5*(11), 396–406, doi:10.4319/lom.2007.5.396.

Kitchen, J. C., and J. R. V. Zaneveld (1992), A three-layered sphere model of the optical properties of phytoplankton, *Limnology and Oceanography*, *37*(8), 1680–1690, doi:10.4319/lo.1992.37.8.1680.

Kiørboe, T. (1993), Turbulence, Phytoplankton Cell Size, and the Structure of Pelagic Food Webs, in *Advances in Marine Biology*, vol. 29, pp. 1–72, Elsevier, doi:10.1016/S0065-2881(08)60129-7.

Kleiber, M. (1932), Body size and metabolism, *Hilgardia*, *6*(11), 315–353, doi:10.3733/hilg.v06n11p315.

Koenka, I. J., J. Sáiz, and P. C. Hauser (2014), Instrumentino: An open-source modular Python framework for controlling Arduino based experimental instruments, *Computer Physics Communications*, *185*(10), 2724–2729, doi:10.1016/j.cpc.2014.06.007.

Koenka, I. J., J. Sáiz, and P. C. Hauser (2015), <I>Instrumentino</I>: An Open-Source Software for Scientific Instruments, *CHIMIA International Journal for Chemistry*, *69*(4), 172–175, doi:10.2533/chimia.2015.172.

Kosobokova, K. N., and R. R. Hopcroft (2010), Diversity and vertical distribution of mesozooplankton in the Arctic's Canada Basin, *Deep Sea Research Part II: Topical Studies in Oceanography*, *57*(1-2), 96–110, doi:10.1016/j.dsr2.2009.08.009.

Kosobokova, K. N., R. R. Hopcroft, and H.-J. Hirche (2011), Patterns of zooplankton diversity through the depths of the Arctic's central basins, *Marine Biodiversity*, *41*(1), 29–50, doi:10.1007/s12526-010-0057-9.

Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2009), Retrieval of the particle size distribution from satellite ocean color observations, *Journal of Geophysical Research*, *114*(C9), C09,015, doi:10.1029/2009JC005303.

Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2010), Global variability of phytoplankton functional types from space: assessment via the particle size distribution, *Biogeosciences Discussions*, *7*(3), 4295–4340, doi:10.5194/bgd-7-4295-2010.

Laney, S. R., and H. M. Sosik (2014), Phytoplankton assemblage structure in and around a massive under-ice bloom in the Chukchi Sea, *Deep Sea Research Part II: Topical Studies in Oceanography*, *105*, 30–41, doi:10.1016/j.dsr2.2014.03.012.

Li, W. K. W., F. A. McLaughlin, C. Lovejoy, and E. C. Carmack (2009), Smallest Algae Thrive As the Arctic Ocean Freshens, *Science*, *326*(5952), 539–539, doi:10.1126/science.1179798.

Litchman, E., and C. A. Klausmeier (2008), Trait-Based Community Ecology of Phytoplankton, *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 615–639, doi:10.1146/annurev.ecolsys.39.110707.173549.

Litchman, E., C. A. Klausmeier, O. M. Schofield, and P. G. Falkowski (2007), The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level, *Ecology Letters*, *10*(12), 1170–1181, doi:10.1111/j.1461-0248.2007.01117.x.

Llort, J., C. Langlais, R. Matear, S. Moreau, A. Lenton, and P. G. Strutton (2018), Evaluating Southern Ocean Carbon Eddy-Pump From Biogeochemical-Argo Floats, *Journal of Geophysical Research: Oceans*, *123*(2), 971–984, doi:10.1002/2017JC012861.

Loisel, H., J.-M. Nicolas, A. Sciandra, D. Stramski, and A. Poteau (2006), Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean, *Journal of Geophysical Research*, *111*(C9), C09,024, doi:10.1029/2005JC003367.

Lombard, F., E. Boss, A. M. Waite, M. Vogt, J. Uitz, L. Stemmann, H. M. Sosik, J. Schulz, J.-B. Romagnan, M. Picheral, J. Pearlman, M. D. Ohman, B. Niehoff, K. O. Möller, P. Miloslavich, A. Lara-Lpez, R. Kudela, R. M. Lopes, R. Kiko, L. Karp-Boss, J. S. Jaffe, M. H. Iversen, J.-O. Irisson, K. Fennel, H. Hauss, L. Guidi, G. Gorsky, S. L. C. Giering, P. Gaube, S. Gallager, G. Dubelaar, R. K. Cowen, F. Carlotti, C. Briseño-Avena, L. Berline, K. Benoit-Bird, N. Bax, S. Batten, S. D. Ayata, L. F. Artigas, and W. Appeltans (2019), Globally Consistent Quantitative Observations of Planktonic Ecosystems, *Frontiers in Marine Science*, *6*, 196, doi:10.3389/fmars.2019.00196.

Longhurst, A. R. (2007), *Ecological geography of the sea*, 2nd ed ed., Academic Press, Amsterdam ; Boston, MA, oCLC: 64336057.

Malkiel, E., O. Alquaddoomi, and J. Katz (1999), Measurements of Plankton Distribution in the Ocean Using Submersible Holography, *Measurement Science and Technology*, *10*(12), 1142–1152, doi:10.1088/0957-0233/10/12/305.

Marañón, E. (2015), Cell Size as a Key Determinant of Phytoplankton Metabolism and Community Structure, *Annual Review of Marine Science*, *7*(1), 241–264, doi:10.1146/annurev-marine-010814-015955.

Marcos, J. R. Seymour, M. Luhar, W. M. Durham, J. G. Mitchell, A. Macke, and R. Stocker (2011), Microbial alignment in flow changes ocean light climate, *Proceedings of the National Academy of Sciences*, *108*(10), 3860–3864, doi:10.1073/pnas.1014576108.

Margalef, R. (1978), Life-forms of phytoplankton as survival alternatives in an unstable environment, *Oceanologica acta*, *1*(4), 493–509, publisher: Gauthier-Villars.

MathWorks (2019), MATLAB Statistics and Machine Learning Toolbox.

McDougall, P., Trevor J. ; Barker (2011), Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox, *SCOR/IAPSO WG127*, p. 28, doi:SCOR/IAPSOWG127.

McLeod, L. E., and M. J. Costello (2017), Light traps for sampling marine biodiversity, *Helgoland Marine Research*, *71*(1), 2, doi:10.1186/s10152-017-0483-1.

Medwin, H. (1975), Speed of sound in water: A simple equation for realistic parameters, *The Journal of the Acoustical Society of America*, *58*(6), 1318–1319, doi:10.1121/1.380790.

Meyer, R. A. (1979), Light scattering from biological cells: dependence of backscatter radiation on membrane thickness and refractive index, *Applied Optics*, *18*(5), 585, doi:10.1364/AO.18.000585.

Mie, G. (1908), Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen, *Annalen der Physik*, *330*(3), 377–445, doi:10.1002/andp.19083300302.

Miller, C. B., and P. A. Wheeler (2012), Population biology of zooplankton, in *Biological Oceanography*, chap. 8, pp. 158–180, John Wiley & Sons.

Moberg, E. A., and H. M. Sosik (2012), Distance maps to estimate cell volume from two-dimensional plankton images: Distance map cell volume algorithm, *Limnology and Oceanography: Methods*, *10*(4), 278–288, doi:10.4319/lom.2012.10.278.

Morel, A., and Y.-H. Ahn (1991), Optics of heterotrophic nanoflagellates and ciliates: A tentative assessment of their scattering role in oceanic waters compared to those of bacterial and algal cells, *Journal of Marine Research*, *49*(1), 177–202, doi:10.1357/002224091784968639.

Morel, A., and A. Bricaud (1986), Inherent Optical Properties of Algal Cells Including Picoplankton: Theoretical and Experimental Results, p. 39.

Morel, A., Y.-H. Ahn, F. Partensky, D. Vaulot, and H. Claustre (1993), Prochlorococcus and Synechococcus: A comparative study of their optical properties in relation to their size and pigmentation, *Journal of Marine Research*, *51*(3), 617–649, doi:10.1357/0022240933223963.

Morel, A., Y. Huot, B. Gentili, P. J. Werdell, S. B. Hooker, and B. a. Franz (2007), Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach, *Remote Sensing of Environment*, *111*(1), 69–88, doi:10.1016/j.rse.2007.03.012.

Morán, X. A. G., A. López-Urrutia, A. Calvo-Díaz, and W. K. W. Li (2010), Increasing importance of small phytoplankton in a warmer ocean, *Global Change Biology*, *16*(3), 1137–1144, doi:10.1111/j.1365-2486.2009.01960.x.

Mouw, C. B., N. J. Hardman-Mountford, S. Alvain, A. Bracher, R. J. W. Brewin, A. Bricaud, A. M. Ciotti, E. Devred, A. Fujiwara, T. Hirata, T. Hirawake, T. S. Kostadinov, S. Roy, and J. Uitz (2017), A Consumer's Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean, *Frontiers in Marine Science*, *4*, doi:10.3389/fmars.2017.00041.

Nayak, A. R., M. N. McFarland, J. M. Sullivan, and M. S. Twardowski (2018), Evidence for ubiquitous preferential particle orientation in representative oceanic shear flows: Nonrandom particle orientation in ocean, *Limnology and Oceanography*, *63*(1), 122–143, doi:10.1002/lno.10618.

Ohman, M. D., and J.-B. Romagnan (2016), Nonlinear effects of body size and optical attenuation on Diel Vertical Migration by zooplankton, *Limnology and Oceanography*, *61*(2), 765–770, doi:10.1002/lno.10251.

Ohman, M. D., R. E. Davis, J. T. Sherman, K. R. Grindley, B. M. Whitmore, C. F. Nickels, and J. S. Ellen (2019), Zooglider: An autonomous vehicle for optical and acoustic sensing of zooplankton, *Limnology and Oceanography: Methods*, *17*(1), 69–86, doi:10.1002/lom3.10301.

Olson, R. J., S. W. Chisholm, E. R. Zettler, and E. V. Armbrust (1990), Pigments, size, and distributions of Synechococcus in the North Atlantic and Pacific Oceans, *Limnology and Oceanography*, *35*(1), 45–58, doi:10.4319/lo.1990.35.1.0045.

Ona, E., O. R. Godø, N. O. Handegard, V. Hjellvik, R. Patel, and G. Pedersen (2007), Silent research vessels are not quiet, *The Journal of the Acoustical Society of America*, *121*(4), EL145–EL150, doi:10.1121/1.2710741.

Organelli, E., G. Dall'Olmo, R. J. W. Brewin, G. A. Tarran, E. Boss, and A. Bricaud (2018), The open-ocean missing backscattering is in the structural complexity of particles, *Nature Communications*, *9*(1), 5439, doi:10.1038/s41467-018-07814-6.

Peña, M. (2019), Mesopelagic fish avoidance from the vessel dynamic positioning system, *ICES Journal of Marine Science*, *76*(3), 734–742, doi:10.1093/icesjms/fsy157.

Picheral, M., L. Guidi, L. Stemmann, D. M. Karl, G. Iddaoud, and G. Gorsky (2010), The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton, *Limnology and Oceanography: Methods*, *8*(9), 462–473, doi:10.4319/lom.2010.8.462.

Picheral, M., S. Colin, and J.-O. Irisson (2017), EcoTaxa, a tool for the taxonomic classification of images.

Pomeroy, L. R. (1974), The Ocean's Food Web, A Changing Paradigm, *BioScience*, *24*(9), 499–504, doi:10.2307/1296885.

Rasse, R., G. Dall'Olmo, J. Graff, T. K. Westberry, V. van Dongen-Vogels, and M. J. Behrenfeld (2017), Evaluating Optical Proxies of Particulate Organic Carbon across the Surface Atlantic Ocean, *Frontiers in Marine Science*, *4*, 367, doi:10.3389/fmars.2017.00367.

Reda, I., and A. Andreas (2004), Solar position algorithm for solar radiation applications, *Solar Energy*, *76*(5), 577–589, doi:10.1016/j.solener.2003.12.003.

Roesler, C., J. Uitz, H. Claustre, E. Boss, X. Xing, E. Organelli, N. Briggs, A. Bricaud, C. Schmechtig, A. Poteau, F. D'Ortenzio, J. Ras, S. Drapeau, N. Haëntjens, and M. Barbieux (2017), Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors, *Limnology and Oceanography: Methods*, *15*(6), 572–585, doi:10.1002/lom3.10185.

Rohart, F., B. Gautier, A. Singh, and K.-A. Lê Cao (2017), mixOmics: An R package for 'omics feature selection and multiple data integration, *PLOS Computational Biology*, *13*(11), e1005,752, doi:10.1371/journal.pcbi.1005752.

Rousseeuw, P. J., and C. Croux (1993), Alternatives to the Median Absolute Deviation, *Journal of the American Statistical Association*, *88*(424), 1273–1283.

Roy, S., S. Sathyendranath, and T. Platt (2011), Retrieval of phytoplankton size from bio-optical measurements: theory and applications, *Journal of The Royal Society Interface*, *8*(58), 650–660, doi:10.1098/rsif.2010.0503.

Roy, S., S. Sathyendranath, H. Bouman, and T. Platt (2013), The global distribution of phytoplankton size spectrum and size classes from their light-absorption spectra derived from satellite data, *Remote Sensing of Environment*, *139*, 185–197, doi:10.1016/j.rse.2013.08.004.

Ryan, T. E., R. A. Downie, R. J. Kloser, and G. Keith (2015), Reducing bias due to noise and attenuation in open-ocean echo integration data, *ICES Journal of Marine Science: Journal du Conseil*, *72*(8), 2482–2493, doi:10.1093/icesjms/fsv121.

Schmechtig, C., A. Poteau, H. Claustre, F. D'Ortenzio, and E. Boss (2015), Processing Bio-Argo chlorophyll-a concentration at the DAC Level, *Tech. rep.*, doi:10.13155/39468.

Schmechtig, C., E. Organelli, A. Poteau, H. Claustre, and F. D'Ortenzio (2017), Processing BGC-Argo CDOM concentration at the DAC level, *Tech. rep.*, doi:http://dx.doi.org/10.13155/54541.

Schmechtig, C., A. Poteau, H. Claustre, F. D'Ortenzio, G. Dall'Olmo, and E. Boss (2018), Processing BGC-Argo particle backscattering at the DAC level, *Tech. rep.*, doi:http://dx.doi.org/10.13155/39459.

Scott, D. W. (1979), On optimal and data-based histograms, *Biometrika*, *66*(3), 605–610, doi:10.1093/biomet/66.3.605.

Shannon, C. E. (1949), Communication in the Presence of Noise, *Proceedings of the IRE*, *37*(1), 10–21.

Sheldon, R. W., A. Prakash, and W. H. Sutcliffe (1972), The Size Distribution of Particles in the Ocean, *Limnology and Oceanography*, *17*(3), 327–340, doi:10.4319/lo.1972.17.3.0327.

Sieburth, J. M., V. Smetacek, and J. Lenz (1978), Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1, *Limnology and Oceanography*, *23*(6), 1256–1263, doi:10.4319/lo.1978.23.6.1256.

Slade, W. H., and E. Boss (2015), Spectral attenuation and backscattering as indicators of average particle size, *Applied Optics*, *54*(24), 7264, doi:10.1364/AO.54.007264.

Slade, W. H., E. Boss, G. Dall'Olmo, M. R. Langner, J. Loftin, M. J. Behrenfeld, C. Roesler, and T. K. Westberry (2010), Underway and Moored Methods for Improving Accuracy in Measurement of Spectral Particulate Absorption and Attenuation, *Journal of Atmospheric and Oceanic Technology*, *27*(10), 1733–1746, doi:10.1175/2010JTECHO755.1.

Sokolova, M., and G. Lapalme (2009), A systematic analysis of performance measures for classification tasks, *Information Processing & Management*, *45*(4), 427–437, doi:10.1016/j.ipm.2009.03.002.

Sommer, S. A., L. Van Woudenberg, P. H. Lenz, G. Cepeda, and E. Goetze (2017), Vertical gradients in species richness and community composition across the twilight zone in the North Pacific Subtropical Gyre, *Molecular Ecology*, *26*(21), 6136–6156, doi:10.1111/mec.14286.

Sosik, H. M., and R. J. Olson (2007), Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry, *Limnology and Oceanography: Methods*, *5*(6), 204–216, doi:10.4319/lom.2007.5.204.

Stedmon, C. A., and R. M. Cory (2014), Biological Origins and Fate of Fluorescent Dissolved Organic Matter in Aquatic Environments, in *Aquatic Organic Matter Fluorescence*, edited by P. Coble, J. Lead, A. Baker, D. M. Reynolds, and R. G. Spencer, pp. 278–300, Cambridge University Press, Cambridge, doi:10.1017/CBO9781139045452.013.

Steinberg, D. K., and M. R. Landry (2017), Zooplankton and the Ocean Carbon Cycle, *Annual Review of Marine Science*, *9*(1), 413–444, doi:10.1146/annurev-marine-010814-015924.

Steinberg, D. K., J. S. Cope, S. E. Wilson, and T. Kobari (2008a), A comparison of mesopelagic mesozooplankton community structure in the subtropical and subarctic North Pacific Ocean, *Deep-Sea Research Part II: Topical Studies in Oceanography*, *55*(14-15), 1615–1635, doi:10.1016/j.dsr2.2008.04.025.

Steinberg, D. K., B. A. S. Van Mooy, K. O. Buesseler, P. W. Boyd, T. Kobari, and D. M. Karl (2008b), Bacterial vs. zooplankton control of sinking particle flux in the ocean's twilight zone, *Limnology and Oceanography*, *53*(4), 1327–1338, doi:10.4319/lo.2008.53.4.1327.

Stramski, D., and D. A. Kiefer (1991), Light scattering by microorganisms in the open ocean, *Progress in Oceanography*, *28*(4), 343–383, doi:10.1016/0079-6611(91)90032-H.

Stramski, D., A. Bricaud, and A. Morel (2001), Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community, *Applied Optics*, *40*(18), 2929, doi:10.1364/AO.40.002929.

Stramski, D., E. Boss, D. Bogucki, and K. J. Voss (2004), The role of seawater constituents in light backscattering in the ocean, *Progress in Oceanography*, *61*(1), 27–56, doi:10.1016/j.pocean.2004.07.001.

Sturges, H. A. (1926), The Choice of a Class Interval, *Journal of the American Statistical Association*, *21*(153), 65– 66.

Sullivan, J. M., M. S. Twardowski, J. Ronald, V. Zaneveld, and C. C. Moore (2013), Measuring optical backscattering in water, in *Light Scattering Reviews 7*, edited by A. A. Kokhanovsky, no. ISBN 978-3-642-10335-3 in Springer Praxis Books, pp. 189–224, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-642-21907-8{_}6.

Talapatra, S., J. Hong, M. McFarland, A. Nayak, C. Zhang, J. Katz, J. Sullivan, M. Twardowski, J. Rines, and P. Donaghay (2013), Characterization of biophysical interactions in the water column using in situ digital holography, *Marine Ecology Progress Series*, *473*, 29–51, doi:10.3354/meps10049.

Tanaka, M., A. Genin, R. M. Lopes, J. R. Strickler, and H. Yamazaki (2019), Biased measurements by stationary turbidity-fluorescence instruments due to phototactic zooplankton behavior, *Limnology and Oceanography: Methods*, p. lom3.10328, doi:10.1002/lom3.10328.

Traykovski, P., R. J. Latter, and J. D. Irish (1999), A laboratory evaluation of the laser in situ scattering and transmissometery instrument using natural sediments, *Marine Geology*, *159*(1-4), 355–367, doi:10.1016/S0025-3227(98)00196-0.

Tremblay, G., C. Belzile, M. Gosselin, M. Poulin, S. Roy, and J. Tremblay (2009), Late summer phytoplankton distribution along a 3500 km transect in Canadian Arctic waters: strong numerical dominance by picoeukaryotes, *Aquatic Microbial Ecology*, *54*, 55–70, doi:10.3354/ame01257.

Twardowski, M. S., E. Boss, J. B. Macdonald, W. S. Pegau, A. H. Barnard, and J. R. V. Zaneveld (2001), A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters, *Journal of Geophysical Research: Oceans*, *106*(C7), 14,129–14,142, doi:10.1029/2000JC000404.

Twardowski, M. S., H. Claustre, S. A. Freeman, D. Stramski, and Y. Huot (2007), Optical backscattering properties of the "clearest" natural waters, *Biogeosciences*, *4*(6), 1041–1058, doi:10.5194/bg-4-1041-2007.

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, *Journal of Geophysical Research*, *111*(C8), doi:10.1029/2005JC003207.

Vaillancourt, R. D. (2004), Light backscattering properties of marine phytoplankton: relationships to cell size, chemical composition and taxonomy, *Journal of Plankton Research*, *26*(2), 191–212, doi:10.1093/plankt/fbh012.

van de Hulst, H. C. (1957), *Light Scattering by Small Particles*, John Wiley & Sons, In.c, N.Y.

Van Heukelem, L., and C. S. Thomas (2001), Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments, *Journal of Chromatography A*, *910*(1), 31–49, doi:10.1016/S0378-4347(00)00603-4.

Vickers, G., and D. Brown (2001), The distribution of projected area and perimeter of convex, solid particles, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *457*(2006), 283–306, doi:10.1098/rspa.2000.0667.

Vinogradov, M. (1970), *Vertical Distribution of the Oceanic Zooplankton.*

Vinogradov, M. (1997), Some Problems of Vertical Distribution of Meso- and Macroplankton in the Ocean, *Advances in Marine Biology*, *32*, 1–92, doi:10.1016/S0065-2881(08)60015-2.

Volten, H., J. F. D. Haan, J. W. Hovenier, R. Schreurs, W. Vassen, a. G. Dekker, H. J. Hoogenboom, F. Charlton, and R. Wouts (1998), Laboratory measurements of angular distributions of light scattered by phytoplankton and silt, *Limnology and Oceanography*, *43*, 1180–1197, doi:10.4319/lo.1998.43.6.1180.

Waterbury, J. B., S. W. Watson, R. R. L. Guillard, and L. E. Brand (1979), Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium, *Nature*, *277*(5694), 293–294, doi:10.1038/277293a0.

Waterbury, J. B., J. M. Willey, D. G. Franks, F. W. Valois, and S. W. Watson (1985), A Cyanobacterium Capable of Swimming Motility, *Science*, *230*(4721), 74–76.

Weithoff, G., and B. E. Beisner (2019), Measures and Approaches in Trait-Based Phytoplankton Community Ecology – From Freshwater to Marine Ecosystems, *Frontiers in Marine Science*, *6*, 40, doi:10.3389/fmars.2019.00040.

Whitmire, A. L., W. S. Pegau, L. Karp-Boss, E. Boss, and T. J. Cowles (2010), Spectral backscattering properties of marine phytoplankton cultures, *Optics Express*, *18*(14), 15,073–15,093, doi:10.1029/2003RG000148.D.

Xing, X., H. Claustre, E. Boss, C. Roesler, E. Organelli, A. Poteau, M. Barbieux, and F. D'Ortenzio (2017), Correction of profiles of in-situ chlorophyll fluorometry for the contribution of fluorescence originating from non-algal matter, *Limnology and Oceanography: Methods*, *15*(1), 80–93, doi:10.1002/lom3.10144.

York, D. (1966), Least-Squares Fitting of a Straight Line, *Canadian Journal of Physics*, *44*, 1079–1086.

Zhang, X., M. Lewis, and B. Johnson (1998), Influence of bubbles on scattering of light in the ocean, *Applied Optics*, *37*(27), 6525, doi:10.1364/AO.37.006525.

Zhang, X., M. Twardowski, and M. Lewis (2011), Retrieving composition and sizes of oceanic particle subpopulations from the volume scattering function, *Applied Optics*, *50*(9), 1240, doi:10.1364/AO.50.001240.

Zheng, G., and D. Stramski (2013), A model based on stacked-constraints approach for partitioning the light absorption coefficient of seawater into phytoplankton and non-phytoplankton components: Partitioning Model of Light Absorption Coefficient, *Journal of Geophysical Research: Oceans*, *118*(4), 2155–2174, doi:10.1002/jgrc.20115.

APPENDIX A

SPIKE LAYER DETECTION ALGORITHM

A.1 Algorithm Description

The algorithm to detect spike layers consists of two major steps: first finding individual spikes in profiles and secondly identifying if the spikes form layers. The algorithm is applied only on sections of the profile with a local sampling depth resolution greater than 1 observation/5 dBar (even if no spike layers were found in profiles with a sampling resolution lower than $\sim$ 1 observation/2 dBar, Figure B.1).

A 15 point Hampel filter (similar to median filter except it only affects outliers) is applied recursively to x, the profile of the optical property of interest (e.g., FDOM, b_{bp}, or fchl), to smooth the profile ($x_{filtered}$). Spikes (x_{spikes}) are considered if they meet the following criteria:

$$x - x_{filtered} > x_{err} \tag{A.1}$$

with x_{err} a threshold defining the minimum intensity of the spikes, this threshold is specific to the property of interest (provided in A.1).

Spikes at depth >10 dBar are clustered together using hierarchical clustering with a euclidean distance cut off of 50 dBar (function cluster from Matlab *MathWorks*, 2019). For each cluster of spikes a set of features are computed and then used for the classification of the profiles: number of spikes (n), median pressure (p), thickness (t), spike density ($d = n/t$), and median normalized spike intensity ($i = median((x_{spike} - min(x))/median(x - min(x)))$. The clusters of spikes meeting the following criteria are considered a layer of spikes potentially caused by mesopelagic organism:

$$\begin{cases} n \geq N \\ T_{min} \leq t \leq T_{max} \\ i \geq I \end{cases} \tag{A.2}$$

With N the minimum number of spike per cluster (set to 3 for this study), T_{min} and T_{max} the minimum and maximum thickness of the spike cluster, and I the normalized intensity threshold of the spike cluster. The value of these parameters are given in Table A.1. A Matlab implementation of this spike layer detection algorithm is available at `https://github.com/OceanOptics/FloatSpikeAnalysis/`.

A.2 Algorithm Performance

The performances of the spike layer detection algorithm was assessed for both FDOM, b_{bp}, and fchl channels with commonly used metrics: accuracy, precision (P), recall (R), and F1-Score (*Sokolova and Lapalme*, 2009). The validation dataset consists of 6889 profiles manually annotated: presence of spike layer (n=127) and absence of spike layer (n=5648), which are the only one used here. Two additional categories, which consist of dubious (n=72) and unusable (n=1042) were ignored. Dubious consists of profiles that could not be classified as present or absent, and unusable consists of shallow profiles (<100 m), under-sampled profiles (<50 observations), or unrealistic profiles (e.g. many negative spikes likely due to dysfunctional sensors).

The assessment of the algorithm (Table A.1) suggest that the algorithm performs well for FDOM profiles. For b_{bp} profiles the precision could be high (>90 %) however many spike layer are missed by the algorithm. For fchl profiles the algorithm is not reliable for detection of mesopelagic organisms which is likely due to the fact that spikes in fchl profiles did not form distinguishable layers and that associated intensities are very small in many cases. The superior performance of the FDOM algorithm resides in the fact that b_{bp} profiles contains additional spikes due to aggregates making it harder to distinguish

zooplankton associated spike layers, while FDOM seems not to be sensitive to those aggregates. A combination of the information contained in both b_{bp} and fchl did not improve the performance of the algorithm with respect to running it for b_{bp} alone. While the statistics of spike layer detection algorithm support using it with no further validation, especially with FDOM profiles, we nonetheless, recommend for ecological studies to validate manually validate all profiles identified as containing spike layers.

Table A.1. Spike layer detection algorithm parameters and associated performance to classify float profiles as containing (present) or not (absence) layers of spikes.

						Presence			Absence		
Channel	x_{err}	I	T_{min}	T_{max}	Accuracy	P	R	F1	P	R	F1
FDOM	3×sMAD*	1.2	3	350	94	68	95	79	99	94	97
FDOM	0.9	1.2	3	350	96	84	86	85	98	98	98
b_{bp}	3×sMAD*	10	10	350	85	40	63	49	95	88	91
b_{bp}	0.00479	10	10	350	94	91	52	66	94	99	97
b_{bp}	0.00281	10	10	350	94	80	65	72	95	98	97
fchl	3×sMAD*	1.2	3	350	82	35	65	46	95	84	89
fchl	0.04	1.2	3	350	83	37	67	48	95	85	90

*sMAD is the scaled median absolute deviation of x over the whole profile as defined by *Rousseeuw and Croux* (1993)

APPENDIX B

ANCILLARY DATA FOR CHAPTER 4

B.1 Float Sampling Resolution

To evaluate the sampling resolution required by the floats to detect spike layers, we first looked at the average distance between observations in each profile of the BGC-Argo database. Only the upper 1000 dBar of the profiles were considered. We then investigated the relative occurrence of profiles with spike layers concerning their sampling frequency. This analysis, presented in Figure B.1, helped to make an informed recommendation on the minimum sampling resolution to detect spike layers in float profiles:
>1 observation/2 dBar. The full spatial coverage of the BGC-Argo floats with backscattering sensors is shown in the map Figure B.1.a, while, the coverage with adequate sampling resolution is shown in the manuscript (Figure 4.3.a).

B.2 NAAMES Acoustics

Three matchups of float profiles and ship-based acoustics of NAAMES (Table B.1) exhibited spike layers in the FDOM and b_{bp} float profiles. The case of September 5-7, 2017 presented in Figure 4.1 and 4.2 is further illustrated with a timeseries of the acoustic at 120 kHz (Figure B.2) which is better suited for observing relatively smaller organisms (1 cm) but limited to the upper 200 m (due to the increased attenuation of sound in that frequency). The timeseries of the acoustic at 120 kHz shows shows similar patterns as the acoustics at 38 kHz for the same period. Besides these matchups of September 5-7 2017 (Campaign 3, Station 2), similar acoustic patterns paired with float spike layers were observed on May 28-31, 2016 (Campaign 2, Station 5) and September 9-11, 2017 (Campaign 3, Station 4) are presented in Figure B.3.

B.3 Experiment with Zooplankton and FDOM Sensors

To test if zooplankton can produce spikes in FDOM signals similar to the in situ observations and to test for phototactic behavior an experiment was set up in 3 phases with sensors similar to the one mounted on the floats (an ECO-FLBBCD and an ECO-FLCDRT).

The experiment was conducted aboard a research vessel at station Papa (north east Pacific), in September 2018, with zooplankton collected $\sim$ 2 hours earlier with a MOCNESS net during daytime. A black tub with a black cover was filled with $\sim$ 80 L of surface seawater and a first set of measurements was taken with both sensors to measure the background signal. Next, $\sim$2 L of highly-concentrated zooplankton was added to the tub and a second set of measurements was taken. Finally, the visible LED of the FLBBCD sensor were covered with black tape such that no visible light was emitted (to avoid inducing a behavioral response towards or away from the sensor) and a third set of measurements was taken (only the FDOM channel is expected to receive a signal). Attention was paid to minimize the tub effect on sensors reading (mainly plastics fluorescence). The tub was kept dark during measurements so that the only possible source of light inside the tub was that of the sensors.

During the first set of measurements (with no zooplankton) no spike were observed with both sensors. The baseline of FDOM slightly increased during the second set of measurements (when zooplankton was present) and spikes similar to the one observed on float profiles were observed.

If the zooplankton are attracted to visible light then we would expect the third set of measurements to have a significantly lower frequency of spikes than the second set of measurements. The frequency of spikes during the third set of measurements was similar to the second one, suggesting that the organisms (mostly *Neocalanus cristatus*, *Vibilia*, *Themisto pacifica*, and *Clausocalanus lividus*) used during this experiment did not manifest an attraction behavior to the ECO sensors. The lack of response could have been induced

by the fact that the zooplankton were recently caught by nets or the species were different than those encountered by the floats with different phototactic sensitivity and behavior.

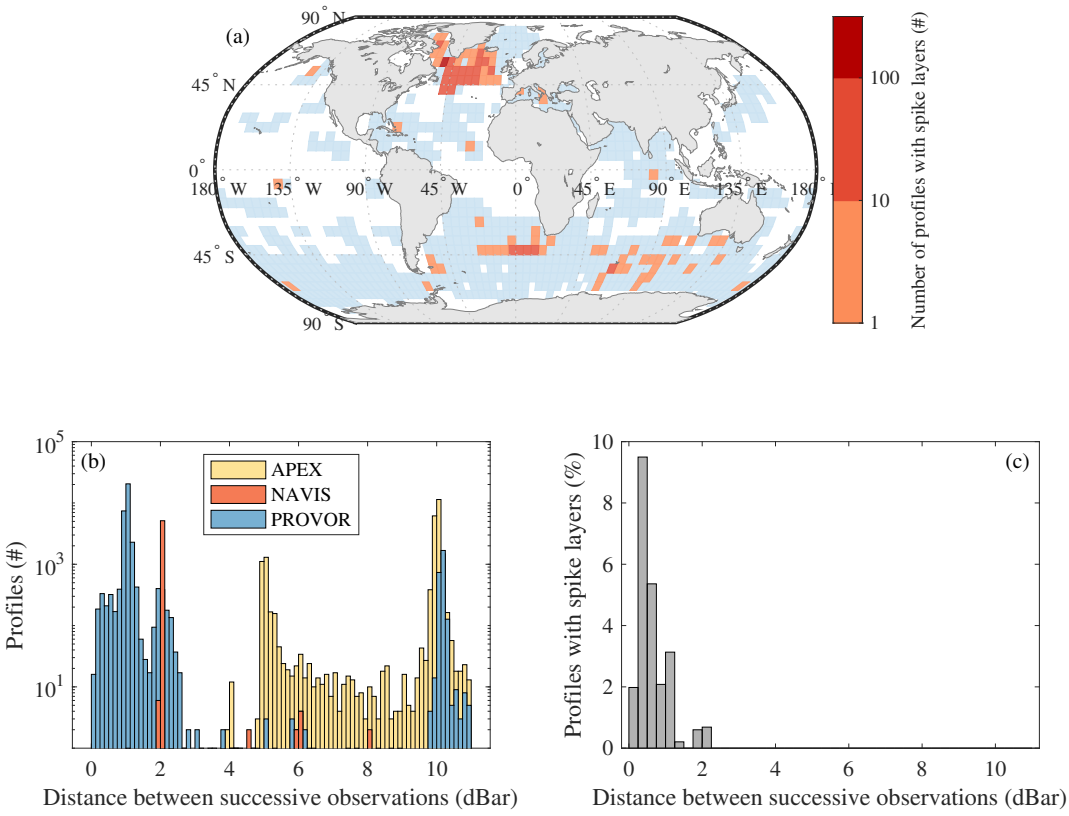

Figure B.1. BGC-Argo Floats Sampling Resolution. (a) Locations of all BGC-Argo float profiles (no restriction on profiling resolution) in a 5×5° grid, blue areas indicate the presence of float profiles and red areas indicate the number of profiles with b_{bp} spike layers. (b) Number of BGC-Argo float profiles as a function of the median distance between successive observations within each profile, colored by platform type: APEX (yellow), NAVIS (red), and PROVOR (blue). Note that 3 % of the profiles with a median observing distance greater than 11 dBar are not shown. (c) Histogram of the number of profiles with spike layers normalized by the total number of profiles as a function of the median distance between successive observations within each profile.

Table B.1. Synthesis of spikes observed with NAAMES BGC-Argo floats collocated with acoustics observations from the R/V Atlantis within eight days. The acoustic backscattering strength S_v is classified as: *(absent)* no evidence of DVM could be visually identified on a 24 hours period, *(weak)* DVM patterns are observed and surface S_v was smaller than -80 dB at night, and *(strong)* DVM patterns are observed and surface S_v was greater than -80 dB at night. The number of spikes per profiles was quantified as follow: *(single)* multiple distinct spikes along the profile, *(collocated)* multiple sets of up to three spikes collocated on the profile, and *(layer)* a layer of spike, more than 5 collocated spikes. Note that the three matchup stations with spike layers are presented in details in Figure 4.1 and A.3. Δz corresponds to the mean distance between the depth of the FDOM spike and the DVM scattering layer, a negative sign means that the spike is deeper than the acoustic layer. All matchups profiles are ascending.

Campaign	Station	WMO	DVM	FDOM Spikes	b_{bp} Spikes	Δz (m)
1	2	5902462	Absent	None	None	—
1	4	5902460	weak	None	None	—
2	1	5903102	weak	None$^{\alpha}$	Single$^{\alpha}$	—
2	2	5902462	weak	Single	Single	200
2	4	6901180*	Strong	—$^{\beta}$	Single$^{\gamma}$	—
2	4	5903101	Strong	None	Single	—
2	5	6901525	Strong	**Layer**	**Layer**	0 ± 0
2	5	5903100	Strong	None$^{\alpha}$	Single$^{\alpha}$	—
3	1	5903109	Strong	None	None	20
3	2	5903108	Strong	**Layer**	**Layer**	30 ± 35
3	3	5903107	Strong	Single	Single	3 ± 5
3	4	5903106	Strong	**Layer**	**Layer**	50 ± 30
3	5	5903105	Strong	Single	Single	-20 ± 23
3	6	5903104	Strong	Collocated	Single	-6 ± 13
3	6	5903103	Strong	Collocated	None	-11 ± 21
4	1	5903109	Strong	None	Single	—
4	2	5903108	Strong	None	None	—
4	2RF	5903108	Strong	None	None	—
4	3	5903107	weak	None	None	—
4	4	5903106	weak	None	None	—

$^{\alpha}$ The float only profiles at dawn.

$^{\beta}$ The float has a channel of $b_{bp}(532)$ in place of FDOM.

$^{\gamma}$ b_{bp} spikes are spread along the entire profile.

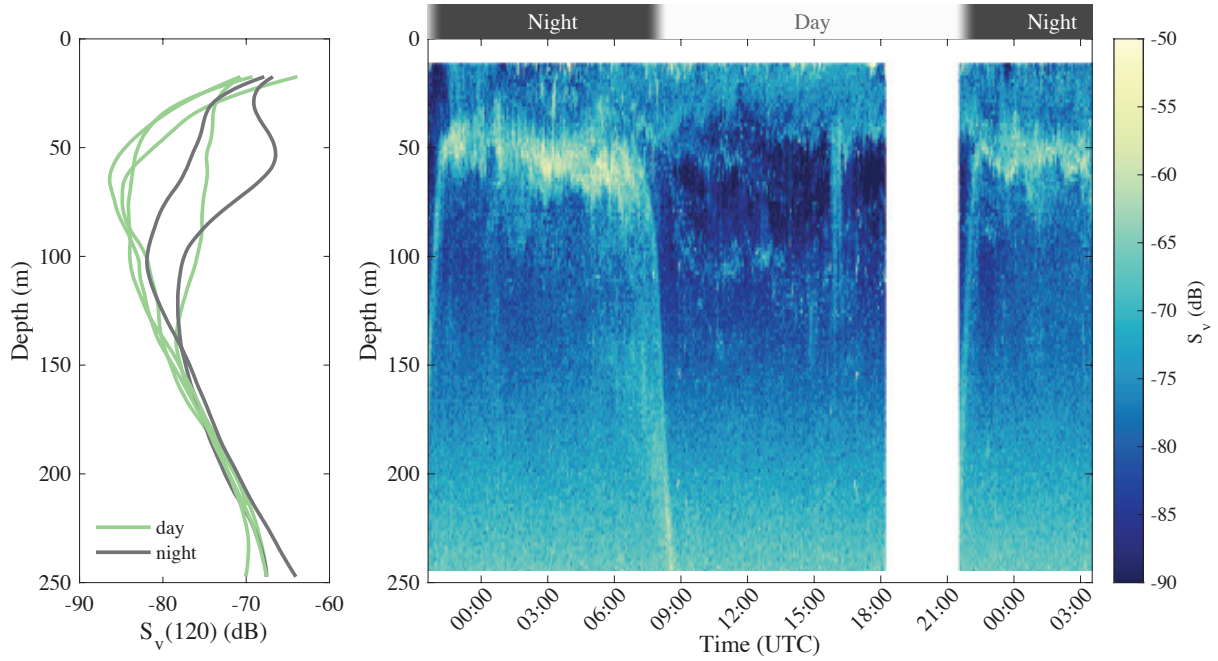

Figure B.2. Profiles *(left)* and timeseries *(right)* of the mean volume backscattering strength at 120 kHz ($S_v(120)$) from the pole-mounted echosounder of the ship at the same time and location of $S_v(38)$ presented in Figure 4.1. Dark lines correspond to night-time profiles and green lines correspond to day-time profiles. The time of the acoustic profiles matches exactly with the time of the float profiles presented in Figure 4.2. Note that the higher profile of $S_v(120)$ during the day time was from 15:00 to 15:30, time during which $S_v(120)$ is temporarily higher. The first 10 m of acoustic data are cropped due to near-field effects. No acoustic backscatter was collected between 18:12 and 21:29.

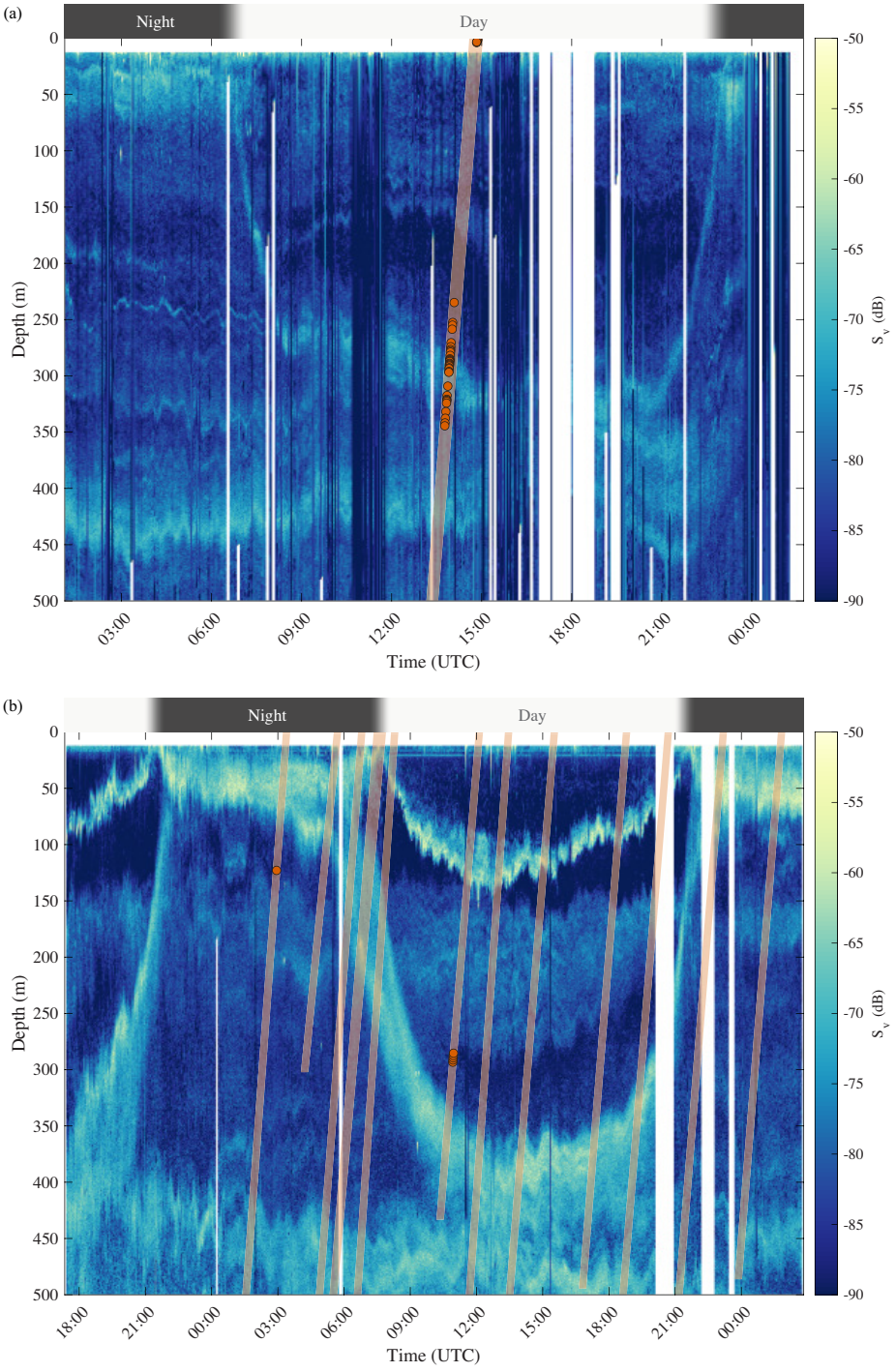

Figure B.3. Time-series of the mean volume backscattering strength (S_v) at 38kHz at the second NAAMES campaign station three *(a)* and the third NAAMES campaign station four *(b)*. The orange lines are the up-casts of the float and orange circles superimposed on the profiles are FDOM spikes. The float profiles are within eight days of the acoustic. The slopes of the lines correspond to the profiling speed (0.08 m/s for the float). The first 10 m of the data are removed to mask near-field effects in the acoustics signal.

BIOGRAPHY OF THE AUTHOR

Nils Haëntjens was born the 9th of July 1991 in Paris, France, to parents Cyril and Anne-Cécile Haëntjens. Nils attended high school in Asnières-sur-Seine, France, graduating in 2009. He then studied at l'Institut Supérieur d'Electronique de Paris (ISEP) where he earned his Master's Degree of Engineer in Embedded Systems in 2014. He spent his second semester of 2014 on the french schooner Tara around the Mediterranean sea. Time during which he met his future advisor, Emmanuel Boss, who led him to Maine and studies in Oceanography. Nils Haëntjens is a candidate for the Doctor of Philosophy degree in Oceanography from the University of Maine in December 2020.